

AD-A138 889

PROBLEM SOLVING AND REASONING(U) PITTSBURGH UNIV PA
LEARNING RESEARCH AND DEVELOPMENT CENTER

1/2

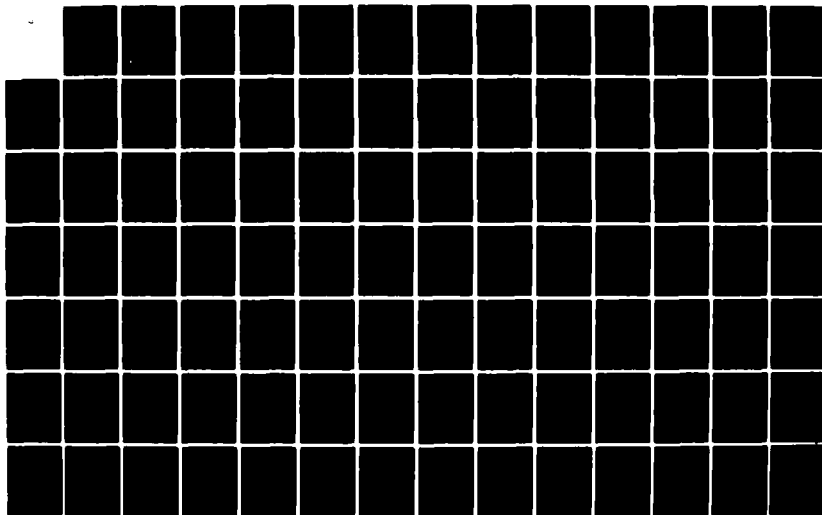
J G GREENO ET AL. FEB 84 UPITT/LRDC/ONR/APS-14

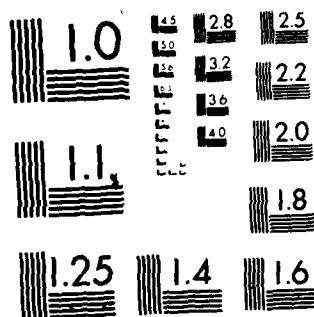
UNCLASSIFIED

N00014-79-C-0215

F/G 5/9

NL





MICROCOPY RESOLUTION TEST CHART
NATIONAL BUREAU OF STANDARDS 1963-A

AD A138889

PROBLEM SOLVING AND REASONING

James G. Greeno
University of Pittsburgh

Herbert A. Simon
Carnegie-Mellon University

February 1984

Technical Report No. UPITT/LRDC/ONR/APS-14

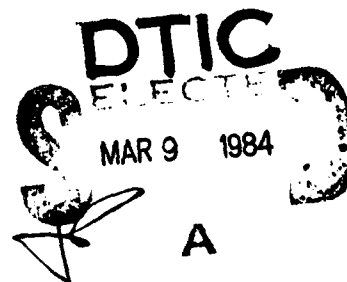
Preparation of this chapter was sponsored by the *Personnel and Training Research Programs, Psychological Sciences Division, Office of Naval Research*, under Contract No. N00014-79-C-0215, Contract Authority Identification Number, NR 667-430.

This report is issued by the Learning Research and Development Center, supported in part as a research and development center by funds from the *National Institute of Education (NIE)*, United States Department of Health, Education, and Welfare.

Reproduction in whole or part is permitted for any purpose of the United States Government.

Approved for public release; distribution unlimited.

DTIC FILE COPY



84 03 08 009

UNCLASSIFIED

SECURITY CLASSIFICATION OF THIS PAGE (When Data Entered)

REPORT DOCUMENTATION PAGE		READ INSTRUCTIONS BEFORE COMPLETING FORM
1. REPORT NUMBER UPITT/LRDC/ONR/APS-14	2. GOVT ACCESSION NO. AD-A138 889	3. RECIPIENT'S CATALOG NUMBER
4. TITLE (and Subtitle) PROBLEM SOLVING AND REASONING		5. TYPE OF REPORT & PERIOD COVERED Technical Report
		6. PERFORMING ORG. REPORT NUMBER
7. AUTHOR(s) James G. Greeno and Herbert A. Simon		8. CONTRACT OR GRANT NUMBER(s) N00014-79-C-0215
9. PERFORMING ORGANIZATION NAME AND ADDRESS University of Pittsburgh Learning Research and Development Center Pittsburgh, PA 15260		10. PROGRAM ELEMENT, PROJECT, TASK AREA & WORK UNIT NUMBERS Prog. Ele.: 61153N Proj. #: RR042-06-0A Task #: NR 667-430
11. CONTROLLING OFFICE NAME AND ADDRESS Personnel and Training Research Program Office of Naval Research (Code 442PTJ) Arlington, VA 22217		12. REPORT DATE February 1984
14. MONITORING AGENCY NAME & ADDRESS (if different from Controlling Office)		13. NUMBER OF PAGES 137
		15. SECURITY CLASS. (of this report) Unclassified
		15a. DECLASSIFICATION/DOWNGRADING SCHEDULE
16. DISTRIBUTION STATEMENT (of this Report) Approved for public release; distribution unlimited.		
17. DISTRIBUTION STATEMENT (of the abstract entered in Block 20, if different from Report)		
18. SUPPLEMENTARY NOTES Preparation of this chapter was also supported by grants from the National Institute of Mental Health and the Alfred P. Sloan Foundation (HAS). This paper is a draft of a chapter to appear in R. C. Atkinson, R. Herrnstein, G. Lindzey, and R. D. Luce (Eds.), <u>Stevens' Handbook of Experimental Psychology</u> , (Revised Edition), New York: John Wiley & Sons.		
19. KEY WORDS (Continue on reverse side if necessary and identify by block number) Problem solving, problem space, search, production system, protocol, weak and strong methods, cognitive models, domain knowledge, planning, design, induction, deduction.		
20. ABSTRACT (Continue on reverse side if necessary and identify by block number) —This long chapter reviews research on problem solving and reasoning; the intended use is as text material for advanced students and others needing a moderately detailed introduction to the topics. The orientation is primarily psychological, with significant attention given to results from artificial intelligence. Major theoretical concepts, such as problem representation, the problem space, strategic knowledge, and problem-solving search, are developed in detail; and major empirical methods such as thinking-aloud protocols, problem-behavior graphs, and use of error patterns and		

DD FORM 1 JAN 73 1473

EDITION OF 1 NOV 65 IS OBSOLETE
5/N 0102-LF-014-6601

UNCLASSIFIED

SECURITY CLASSIFICATION OF THIS PAGE (When Data Entered)

UNCLASSIFIED

SECURITY CLASSIFICATION OF THIS PAGE (When Data Entered)

latencies, are described and illustrated. Sections of the chapter include: Problems with well specified goals and procedures, Problems of design and arrangement, Inductive problem solving, and Evaluation of deductive arguments.

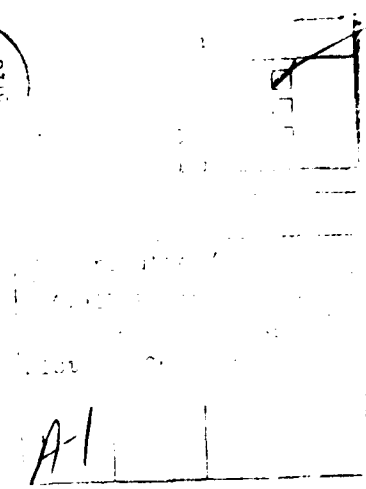
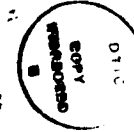
UNCLASSIFIED

SECURITY CLASSIFICATION OF THIS PAGE (When Data Entered)

PROBLEM SOLVING AND REASONING

James G. Greeno
University of Pittsburgh

Herbert A. Simon
Carnegie-Mellon University



ABSTRACT

This long chapter reviews research on problem solving and reasoning; the intended use is as text material for advanced students and others needing a moderately detailed introduction to the topics. The orientation is primarily psychological, with significant attention given to results from artificial intelligence. Major theoretical concepts, such as problem representation, the problem space, strategic knowledge, and problem-solving search, are developed in detail; and major empirical methods such as thinking-aloud protocols, problem-behavior graphs, and use of error patterns and latencies, are described and illustrated. Sections of the chapter include: Problems with well specified goals and procedures, Problems of design and arrangement, Inductive problem solving, and Evaluation of deductive arguments.

This is a draft of a chapter to appear in R. C. Atkinson, R. Herrnstein, G. Lindzey, and R. D. Luce (Eds.), Stevens' Handbook of Experimental Psychology, (Revised Edition). New York: John Wiley & Sons.

PROBLEM SOLVING AND REASONING
James G. Greeno and Herbert A. Simon

I. Introduction

Important advances have been achieved in the 1960's and 1970's in the scientific study of thinking. These advances have resulted from new methods for formulating models of the cognitive processes and structures underlying performance in complex tasks, and the development of experimental methods to test such models. A major accomplishment has been the discovery of general forms of cognitive activity and knowledge that underlie human problem solving and reasoning. This chapter describes a survey of the major theoretical concepts and principles that have been developed, presents some of the evidence that supports these principles, and discusses the empirical and theoretical methods that are used in this domain of scientific study. In this introductory section, we give an overview of the major concepts that will be described in detail in the chapter, and we discuss relations between these concepts and issues that have been investigated previously in experimental psychology. We also discuss some general methodological issues.

I.A. Overview of Concepts

The concepts that have been developed can be discussed conveniently in two groups: hypotheses about the form of cognitive action, and hypotheses about the form of cognitive representation. The hypotheses about cognitive action extend analyses of behavior that were developed in general behavior theory by investigators such as Thorndike (e.g., 1923), Tolman (e.g., 1928); Skinner (e.g., 1938), and Hull (e.g., 1943). The hypotheses about representation extend analyses that were developed by Gestalt psychologists such as Kohler (1929), Duncker (1935/1945), Katona (1940), and Wertheimer (1945/1959). One of the important insights reached in the analysis of problem solving is that hypotheses about these issues of action and representation are complementary, and both are necessary components of a theory of human thought. We will discuss the two groups of concepts in turn in this overview; however, in the sections that follow, hypotheses about action and representation will be integrally related in models of performance in specific tasks.

I.A.1. Form of Cognitive Action. Hypotheses about cognitive action can be considered at two levels: basic action knowledge and strategic knowledge.

A consensus has developed that human knowledge underlying cognitive action can be represented in the form of production rules, a formalism introduced by Post (1943) to represent reasoning in mathematics, and adapted for application to psychology by Newell and Simon (1972). Models in which knowledge for action is represented as a set of production rules are referred to as production systems.

Any theory of performance must include hypotheses about the process of choice whereby individuals select the actions that they perform. A production system provides a framework for expressing hypotheses about this process in specific detail. A production rule (or, more simply, a production) consists of a condition and an action. The condition specifies

a pattern of information that may or may not be present in the situation. The action specifies something that can be performed. The general form of action based on productions is simply: if the condition is true, perform the action.

In a production system, the basic problem of choice among actions is solved by specifying conditions that lead to the selection of each action that can be performed. The condition of each production rule is a pattern of information that the system can recognize. These patterns include features of the external problem situation (the stimulus). They also include information that is generated internally by the problem solver and held in short-term memory. The internal information includes goals that are set during problem solving. It also can include information in memory, such as past attempts to achieve specific goals. Thus, production rules, which represent basic action knowledge, consist of associations between patterns of information and actions. An action is chosen when the individual has a goal with which the action is associated and the external stimulus situation as well as information in memory include features associated with the action.

An important component of a model of cognitive activity is its representation of strategic knowledge. This includes processes for setting goals and adopting general plans or methods in working on a problem. Analyses of general problem-solving strategies have been developed to simulate performance in novel problem situations where the individual has little or no experience. One major analysis is based on a process of means-ends analysis (e.g., Newell & Simon, 1972) in which goals are compared with current states, and actions are selected to reduce differences that are identified. General strategies also include processes for setting subgoals when the current goal cannot be achieved directly. Analyses of strategic knowledge in specific domains also have been developed to simulate performance by problem solvers who have received special training (e.g., Greeno, 1978). Strategic knowledge of experienced problem solvers includes global plans for solving classes of problems and knowledge of subgoals that are useful in classes of problem situations.

The general ideas used in formulating hypotheses about cognitive activity in production systems can be regarded as building upon, rather than negating, the concepts developed and used in general behavior theory, particularly the formulations of Tolman, and the later forms of Hull's theory. Early expositions of behavior theory emphasized the direct relations between stimuli and responses, with rather deliberate inattention to intervening events in the brain. Thorndike (1923) emphasized that actions are chosen because of their associations or bonds with stimulus conditions. In Skinner's (1938) formulation, actions are performed under the "control" of external stimulus features. Tolman (1928), on the other hand, strongly recognized the need to include internal goals and information stored in memory in the determination of response selection. Tolman used such terms as "means-end expectation" and "means-end readiness" in referring to these factors. In Hull's theory, concepts of covert anticipatory responses (1930) and incentive motivation (1952) were used. In discussions of problem solving, Maltzman (1955) and Staats (1966) postulated stimulus-response units at differing levels of generality, and the idea of knowledge about action at different levels is used in more recent formulations of strategic knowledge, especially in hypotheses about

planning.

The concept of a production rule is consistent with these formulations; and behavior theory, even in the terms used by Watson and Skinner, can be expressed as a system of productions (Millenson, 1967). However, as production rules are used in contemporary information processing theory, they make much more explicit than did earlier theories the motivational states and memories of prior experiences that combine with external stimulus conditions to determine choice of a response. Modern production system models of problem solving and similar cognitive processes may be viewed as a (lengthy) extrapolation of Tolman's research program that symmetrizes the roles of external environment (stimulus) and inner environment (motivational states and memory contents) as determinants of response; and that makes far more explicit than earlier formulations were able to exactly how those two sources of information control responses. We characterize the extrapolation as "lengthy" because not only does it postulate that many of the essential components of the stimulus lie in the brain, but also that a large part of the response to a production (or all of it) may be internal -- consisting, for example, of a change in content of short-term or long-term memory. We do not want to underestimate the magnitude of the shift in viewpoint, but we do wish to emphasize that it is a continuous development from the experimental psychology that preceded it, and not a new start. That is presumably what Miller, Galanter, and Pribram (1960) also meant when they described the new approach (half jokingly) as "subjective behaviorism." ("Subjective," of course, referred to the minds of the subjects, not to the scientific methods of the investigators.)

One major difference between recent hypotheses about cognitive activity and those developed in general behavior theory, in addition to the shift to internal events in behavior, is that recent formulations are much more definite and specific. Models have been formulated as production systems with sufficient specificity that they can be expressed as computer programs that simulate actual performance of solving specific problems. To do this, it is not sufficient to postulate the existence of stimulus-response associations and goals, even at differing levels of generality; it is necessary also to formulate hypotheses about just what the stimuli, responses, and goals are. Hypotheses about specific structures of knowledge about actions and goals in the problem domain have to be constructed, and processes have to be designed to recognize specific patterns of information in the task situation that are relevant to solving problems. Hypotheses about strategic knowledge have to specify the conditions in which goals will be set and plans will be adopted.

Again, we prefer to emphasize continuity, rather than discontinuity in this development. Nothing in the new fine-grained mechanisms is antithetical to the grosser level of description of the earlier theories. In fact, important progress has been made in explaining in detail (and sometimes quantitatively) the rich body of experimental data provided within the behavioral scheme (e.g., Simon & Feigenbaum, 1964; Gregg & Simon, 1967). But the impact from achieving this higher level of resolution in our theoretical models and their predictions has led to significantly greater understanding of the psychological processes involved in problem solving and reasoning.

I.A.2. Hypotheses about Representation. Hypotheses about cognitive representations of problems are formulated using the idea of a problem space. The problem space includes an individual's representation of the objects in the problem situation, the goal of the problem, and the actions that can be performed and strategies that can be used in working on the problem. It also includes knowledge of constraints in the problem situation: restrictions on what can be done, as well as limits on the ways in which objects or features of objects can be combined.

In developing hypotheses about representation of problems, much use has been made of concepts developed in analyses of language understanding, including networks of propositions (Quillian, 1968; Kintsch, 1974; Anderson, 1976), procedural representation of concepts (Feigenbaum, 1963; Hunt, Marin & Stone, 1966; Winograd, 1972), and schemata (Schank, 1972; Hayes & Simon, 1974; Norman & Rumelhart, 1975; Schank & Abelson, 1978). Representations of problems differ from those usually postulated for understanding of language because they are constrained to provide information needed for solving the problem. Hypotheses about knowledge used in representing problems include processes for recognizing features that are relevant to actions, strategies, and constraints of the problem domain, and for constructing representations with information that can be used in the cognitive processes of problem solving.

Hypotheses about problem representations have begun to address some of the issues of understanding principles and structure in problem solving that were emphasized by some educational, developmental, and Gestalt psychologists (e.g., Judd, 1908; Kohler, 1929; Brownell, 1935; Duncker, 1935/1945; Katona, 1940; Piaget, 1941/1952; Wertheimer, 1945/1959). As with hypotheses about cognitive activity, current hypotheses about representation are more definite and specific than those that characterized previous discussions. The hypotheses specify cognitive processes and structures that actually construct representations from the texts or other presentations of problem information (e.g., Hayes & Simon, 1974; Larkin, McDermott, Simon & Simon, 1978; Riley, Greeno & Heller, 1983). Hypotheses about understanding of problem structure and general principles include cognitive structures that specify just what is understood about the problem, and how the understanding is achieved (Greeno, 1983; Greeno, Riley & Gelman, in press). Another characteristic of recent discussions is that hypotheses about understanding are coordinated with hypotheses about cognitive activity in problem solving, so the significance of understanding, as well as the specific information that it provides for the problem solver, is made clear.

I.B. Methodology

The use of computer programming languages as formal systems for psychological theory has been a major factor in the development of the concepts and empirical results that we describe in this chapter. The standards that are now common for adequacy of a hypothesis include its expression in a computer program that simulates actual solution of problems — that is, a description of the problem can be given as input for the program, and the program carries out steps that result in the problem's being solved. To meet this standard, the theorist must develop specific hypotheses about many aspects of the psychological process that were previously left unspecified. Representations of specific stimulus

situations must be postulated, including relations among cues that are assumed to provide important information for the subject. Knowledge structures and processes required for comprehension of stimulus situations must be specified, leading to specific forms of information that are assumed to constitute the subject's cognitive representations of the stimuli. Assumptions about knowledge in the subject's memory are specified in detail, including associative structures of information and production rules in which specific actions are associated with specific stimulus conditions. The actions include overt responses as well as internal actions such as setting goals and choosing plans.

To provide evidence for these more detailed hypotheses, more detailed data are required. A major source of these data has been the increased use of thinking-aloud protocols. Protocols provide a more detailed description of behavior, enabling inferences about intermediate steps such as subgoals and attention to specific aspects of the problem. Protocol statements are not treated as introspective descriptions of psychological processes, but rather as overt reports of mental activity that the subject would be aware of in any case, but usually would not announce. Indeed, subjects are instructed to avoid trying to explain their behavior, but only to give reports of things they notice or think about as they are working (cf. Ericsson & Simon, 1980). Statements in protocols provide data to be explained by models that constitute hypotheses about the process, and thus protocol statements have the same status as other detailed observations, such as specific patterns of errors by individuals on sets of problems, latencies of response when information for problems is presented sequentially, or eye fixations during processing of problem information.

I.C. Chapter Contents

The remainder of this chapter is organized in five sections. We have organized the findings and conclusions that we present on the basis of general properties of the cognitive tasks that have been studied.

Section II deals with problems in which a definite goal or solution procedure is specified. Analyses of problems of this kind have been especially important in the development of concepts and methodology, and we have devoted more space to Section II than we have to the other sections. In Section II we develop general theoretical ideas, such as the problem space and heuristic search, that are used without detailed development in later sections. Section II also includes discussion of methodology and empirical evidence in more detail than later sections. Conclusions presented in other sections are based on evidence similar to that discussed in Section II, although space did not permit us to describe the empirical findings as fully in those later discussions.

Examples of problems specified by goals or procedures include logic exercises, where the goal is to derive a specified expression, and arithmetic problems in which a child must perform the steps of subtraction. These problems present a situation and require performance of a sequence of actions that transform the situation. A limited set of problem-solving operators are permitted, restricting the actions that can be performed. The subject's task can be viewed as a search in a space of action sequences, where there are many possible sequences of actions, only a few of which lead to the problem goal and conform to the constraints of the

situation.

In Section III we consider problems of design and arrangement, where goals are specified in terms of general criteria, rather than as definite states or procedures. A familiar example is an anagram problem, where a set of letters is to be arranged into a word. Problems of design differ from the transformation problems discussed in Section II in that constraints are imposed mainly on the solution state, rather than on the actions that can be used in achieving the state. Thus, design problems can be understood as problems of search in a space that contains many possible arrangements of the problem materials, only one or a few of which satisfy the problem criterion.

In Sections IV and V we consider tasks that are often called reasoning rather than problem solving. Section IV takes up problems of induction, and Section V deals with deductive syllogisms. Analyses of processes involved in these tasks reveal that they share basic characteristics of the processes involved in tasks ordinarily considered as problem solving, although they also have some distinctive features. In induction problems, the goal is to identify the structure of a set of materials; the problems require search in a space that contains many possible structural descriptions or rules, most of which are inconsistent with some features of the problem information. In tasks frequently used to study deductive reasoning, problem solvers judge whether arguments that are presented are valid; the process involves an attempt to derive the conclusion from the premises, a search for a sequence of inferential actions just as in problems of transformation.

In Section VI we present some conclusions.

II. Well Specified Problems

In this section we discuss problem solving in relatively well structured situations. First, we consider problems in which a definite goal is specified. The problem solver is given an initial situation or problem state, a set of operators that can be used to change the situation, and a goal state. The task is to find a sequence of actions, restricted to use of the permitted operators, that results in the goal state. In Section II.A we discuss goal-directed problems for which the problem solver has little or no specific knowledge or experience, so that the problem solving depends on using general problem-solving knowledge sometimes called "weak methods." In Section II.B. we discuss problems of the same structure for which individuals have received special training or experience, thus acquiring specific knowledge of the problem domain. Section II.C discusses problems in which procedure is specified, rather than a goal, and correct performance depends on the subject's knowledge of the procedure. In Section II. we discuss representation of problems, including understanding in operations for novel problems and representing word problems in domains for which the individual has received special training.

II.A. General Knowledge for Novel Problems with Specific Goals

A substantial body of research has been conducted on solution of well-structured puzzle-like problems that require relatively little domain-specific knowledge. By hindsight, the research strategy of focusing on such problems has some advantages, even beyond the obvious ones of making the experiments simpler and the data easier to interpret. In difficult problem domains requiring special knowledge, we are likely to learn from our subjects principally what they know and how they have organized and represented their knowledge in memory, because much of an individual's success depends on whether he or she knows the specific principles and procedures of the domain.

In experiments in domains that are relatively free of specialized content and where subjects are relatively naive, we may still find significant differences in behavior from subject to subject and from domain to domain, but we also are likely to discover some of the commonalities of behavior that characterize problem solving, at least by novices, over a wide range of domains. We also are likely to detect the flexible, general-purpose techniques that people fall back on when they do not have special knowledge or methods adapted specifically to the task at hand. These fall-back techniques, often called "weak methods," are the only weapons that are available for attacking truly novel problems. Hence, understanding them should contribute also to an understanding of discovery processes and creative problem solving.

An important general concept in the analysis of problem solving is the problem space, consisting of the problem solver's representation of the materials of the problem along with knowledge that is relevant to the task. The problem space includes a representation of the problem goal and operators that can be used; these may be specified in the problem description or supplied by the problem solver's knowledge. The operators include actions that can be performed and conditions that are required for performance of the actions. The problem space also includes the problem

solver's strategic knowledge, which may include methods previously acquired through experience in the domain, as well as general problem-solving methods.

In this section we discuss tasks in which definite goals are specified in the problem instructions. Subjects solving these problems are usually not experienced in the tasks. The problem-solving operators also are specified in the problem instructions, rather than being known in advance by the problem solvers, and the problem solvers must rely on general problem-solving strategies, that is, on weak methods. The principal methods of this kind employ a general problem-solving heuristic called means-ends analysis, a process that involves comparing the current state with the goal of the problem or a subgoal that the problem solver is trying to achieve, and selecting an operator that can reduce differences between the current state and the goal.

Research has been conducted on several tasks of this general kind. Here we discuss two tasks: proof discovery exercises in logic (Newell & Simon, 1972), and water-jar problems (Atwood & Polson, 1976). The studies that we discuss illustrate use of two empirical methods. Newell and Simon's study of logic proof discovery used detailed analyses of thinking-aloud protocols obtained from a few subjects, with data from a larger group of subjects to check the representativeness of some general features of performance. Atwood and Polson's study of water jar problems used frequencies of responses that occurred during problem solving to evaluate a model of problem solving expressed in quantitative form.

II.A.1. Discovering Proofs in Logic. Discovering proofs for mathematical theorems of one kind or another is a task all of us have faced frequently in school and a few of us in our professional lives. One domain in which theorem proving has been studied extensively is elementary symbolic logic (Moore & Anderson, 1954; Newell, Shaw, & Simon, 1957; Newell & Simon, 1972). The propositional calculus is defined by only two rules of inference and a dozen axioms. In the studies that we discuss, the task was presented as a syntactic game of transforming strings of uninterpreted symbols according to rules given as symbolic formulas. This ensured that subjects could not draw readily on such commonsense knowledge as they may have had of the laws of reasoning. (The studies of syllogistic reasoning that we discuss in Section V directly address the question of subjects' knowledge of formal logical rules.)

Deduction and Induction in Problem Solving. At the outset we must deal with one common misconception about proof-finding tasks. Logic is the science of deductive reasoning from premises to conclusions. A proof is a sequence of expressions starting with axioms (or previously proved expressions) and terminating with the desired theorem; each step of the proof must satisfy the laws of deduction. Its validity can be checked, step by step, by applying those laws systematically.

Finding the proof of a theorem is another matter. We have a known starting point -- the axioms -- and a known goal -- the theorem -- but in most mathematical domains there is no systematic rule for constructing a path from axioms to theorem. That path must be discovered, and the usual method for discovering it is to search for it, the amount of trial and error required depending on how selectively the search is carried out.

Hence, while a proof is an example of a logical deduction, the problem-solving activity involved in searching for a proof is inductive search, as is most interesting problem solving whatever the task domain may be.

The Moore-Anderson Logic Problems. In the logic task designed by Moore and Anderson (1954), subjects were not told that they were discovering proofs in symbolic logic, but were simply instructed to "recode" certain strings of symbols into other, specified, strings, using a given set of transformation rules. The rules were displayed on a sheet of paper, which was available to the subjects at all times. A typical rule (there were twelve, some with subparts) was:

$$A \vee B \rightarrow B \vee A,$$

which was to be interpreted: The expression $A \vee B$ may be transformed into the expression $B \vee A$, where A and B are variables for which any parts of an expression can be substituted. The connectives in such expressions were referred to by the experimenter as "wedge," "dot," "horseshoe," and "tilde," instead of being given their usual interpretations in logic as "or," "and," "implies," and "not." Subjects were run on this task by Carpenter, Moore, Snyder, and Lysansky (1961) at Yale, and by Newell and Simon (1972) at Carnegie Institute of Technology.

Several kinds of data can be obtained in problem-solving tasks of this kind. The times to solution can be recorded, as well as the times for making each successive transformation of an expression. Numbers of correct solutions can be counted, and errors can be classified and analysed.

Thinking-aloud Protocols. The richest data, however, are obtained by instructing subjects to think aloud while solving the problem. The verbal protocols provide a higher temporal density of data than is usually obtained by other methods (except, perhaps, from records of eye movements). Typically, subjects speak at an average rate of about two words per second, although there are of course substantial differences among subjects and from one part of a task to another.

If thinking-aloud data are to be used correctly and effectively to help understand subjects' cognitive processes, answers are needed to several questions, especially: (1) which processes, or what parts of the processes, are verbalized, and (2) to what extent does verbalization alter or in any way affect the problem-solving process itself. A recent extensive review of relevant literature (Ericsson & Simon, 1980) supports three general conclusions. First, subjects mainly verbalize a subset of the symbols that pass through STM as the task is being performed. The verbalizations will be more complete (i.e., will give a fuller record of successive STM contents) if the problem is being solved in terms of verbal symbols than if the STM contents have to be translated from some other modality (i.e., visual images). Second, the process of recognizing some familiar visual or auditory stimulus does not produce any intermediate symbols in STM that can be reported; only the result of the recognition process can be reported. Third, in most problem-solving tasks, the cognitive processes are the same in the thinking-aloud as in the silent condition. Moreover, in general, the speed of task performance is neither increased nor decreased by the instructions to think aloud.

The protocols under discussion here are those produced by subjects concurrently with their performing the cognitive task. In using retrospective protocols as data, additional factors must be taken into consideration. First, only such information can be reported retrospectively as has been transferred to LTM and retained there. Second, unless the instructions call for recall of specific events, subjects may engage, in a variety of ways, in active reconstruction of the event or process that is being probed. Hence, retrospective protocols must be interpreted in the light of what we know about the laws of memory and forgetting (Bartlett, 1932; Nisbett & Wilson, 1977).

The most detailed analysis of problem-solving protocols calls for reconstructing from them the successive cognitive states of subjects as they work toward the problem solution. "Cognitive state" means what the subject knows or has found out about the problem up to the time of the protocol fragment being examined, along with information, such as subgoals and evaluations, that has been generated by the subject from decisions and judgments. Typically, in tasks like the logic theorem proving task, subjects verbalize the symbolic expressions they produce and those they are actively considering, the operators they are applying to transform expressions, and often the goals they are trying to attain (e.g., the final theorem or expressions they think would bring them closer to it) (Newell & Simon, 1972). As they proceed, subjects often evaluate their progress and the suitability of steps they have just taken.

From such protocol statements we can usually reconstruct the problem space in which a subject is operating. Recall that a problem space includes a subject's representation of the problem situation, the goal, problem-solving operators, constraints, and strategic knowledge. More formally, a problem space is defined by a set of symbol structures, corresponding to the cognitive states that can be generated as the subject works on the task, and a set of cognitive operators, information processes that produce new cognitive states from existing ones. The problem-solving efforts of a subject may be described as searches through a problem space, from one cognitive state to another, until the solution (a particular cognitive state) is found or the search is abandoned.

Given a description of the problem space, inferred from a protocol, a search tree, called a Problem Behavior Graph (PBG), can be constructed to represent the course of the subject's search. The size and shape of the PBG will disclose the extent of the subject's skill and knowledge and the consequent selectivity he is able to achieve. Given the PBG, in turn, the experimenter can undertake to construct a simulation program for a computer which, if given the same problem, would generate the same PBG as that generated by the subject.

The accuracy of fit of the simulation program to the strategy that guides a subject's behavior can be judged by comparing the program's trace step by step with the problem-solving protocol. Formal methods for judging goodness of fit in a statistical sense are not available, but departures of trace from protocol are easy to detect. These discrepancies then form the basis for modifying the simulation program to fit the protocol better. Except for the fact that the data we are dealing with here are not numerical, the process of fitting a computer program to protocol data is identical in principle with the process of fitting a system of differential

equations to time series data.

A basic problem space for the logic task is one in which the subject's cognitive state is defined by the logic expressions thus far derived from the initial given expression, and by the legal operators for generating new expressions from these. Since the protocol normally discloses both what operators are being applied and what expressions are obtained from the application, there will be a great deal of redundancy in the available information to test the consistency of the interpretation. Many protocols will allow a richer problem space to be inferred -- one in which the subject notes similarities and differences among logic expressions, and chooses his or her next step in terms of them. When the subject's choice of actions is also guided by goals and subgoals, these are also added to the description of the problem space.

Solution Processes. No single strategy, or simulation program based on such a strategy, can be expected to describe the problem-solving behavior of all subjects. However, the behavior of a great many subjects in task domains like logic theorem proving reveals a small number of common mechanisms as central features of the problem-solving process. One of the most important of these is means-ends analysis, first introduced into the problem-solving literature by Duncker (1935/1945). Means-ends analysis requires a problem space rich enough to contain not only logic expressions and operators, but also symbol structures that describe differences between pairs of logic expressions and other symbol structures that describe goals. Thus, a subject operating in such a problem space might say, "I have an expression whose main connective is a horseshoe, and my goal expression has a wedge. Let me look for an operator that will change horseshoe to wedge."

In broadest outline, means-ends analysis can be described by the following set of productions, where S is the present state or expression, G is the goal expression, D is a difference between two expressions, and O is an operator:

- If the goal is to remove difference D between S and G
 - find a relevant operator O
 - and set the goal of applying it.
- If the goal is to apply O to S,
 - and condition C for applying O is unsatisfied
 - set the goal of satisfying C by modifying S.
- If the goal is to apply O to S → make application.
- If there is a difference D between S and G
 - set the goal of removing it.
- If there is no difference between S and G
 - halt and report problem solved.

While the production system displayed here does not describe all the details of the control of search, it provides the main outlines of means-ends analysis. The system seeks to detect a difference between the present position in the problem space and the goal position. Given such a difference, it searches memory for an operator that is relevant for removing the difference. Having found an operator, it attempts to apply it. If all the conditions for operator application are not satisfied, it expresses the discrepancy as a new difference and establishes the goal of reducing it. The scheme operates recursively, and as soon as one difference has been removed, it looks for another. An important component

of the strategy not represented in the productions is the use of memory to store goals that have been tried, so the problem solver can avoid looping through the same cycle of repeated unsuccessful attempts of a goal that cannot be achieved.

A clear distinction can be made between the general strategy of means-ends analysis and domain-specific knowledge that is required for the strategy to be used in solving any particular problem. The general strategy is represented in the productions shown above. To use these productions, a problem solver must be able to represent the state S and the goal G , and identify differences between them. In the domain of logic, states correspond to expressions, and differences involve different letters, different connectives, and different arrangements of letters and connectives. The problem solver also must know what operators can be used, what conditions permit each operator to be applied, and what kinds of difference are removed by use of each operator. In logic, the operators are the rules for transforming expressions. The conditions are patterns that are specified in the rules, and the relevant differences for a rule can be inferred by comparing the two sides of the rule. For example, $A \cdot B \rightarrow A$ requires a pattern in which two subexpressions are connected by a dot, and has the effect of removing a letter or a subexpression, as well as removing the dot. $A \supset B \leftrightarrow \sim A \vee B$ does not remove or add any letters, it can be applied to a pattern with a horseshoe to change the horseshoe to a wedge or vice versa, and it changes the sign of one of the letters or subexpressions.

The general strategy of means-ends analysis has been implemented in a program called the General Problem Solver (GPS) and shown to be sufficient for providing solutions in over a dozen problem domains, including puzzles such as the Tower of Hanoi and tasks such as integral calculus, given appropriate representations of the states, operators, and connections between operators and differences in the specific domains (Ernst & Newell, 1969).

In the experiments conducted with the logic task, subjects were not experienced in the domain. The operators were presented as part of the task instructions, and it is reasonable to expect that subjects had to rely mainly on general problem-solving strategies, rather than having domain-specific knowledge available for the task. If this is correct, and if the subjects' general problem-solving strategies have the properties of GPS, then their performance in the logic task should be similar to that of GPS when it is run on the task. The results were quite positive.

Kinds of Evidence. The hypothesis was evaluated at three levels. First, specific protocols were examined, comparing the statements made by subjects with the steps in solutions by specific versions of GPS. For these simulations, GPS was varied by supplying it with differing priorities of differences. Second, a set of protocols (all those obtained by Newell and Simon on one moderately difficult problem) were coded and each protocol was translated into a Problem-Behavior Graph, showing a succession of cognitive states that was inferred from the statements and problem-solving operators to account for the transitions between states. The state-to-state transitions were classified, and the categories were compared with categories of activity that are performed by GPS. Third, some summary statistics were compiled for Newell and Simon's subjects and

for the subjects run at Yale, involving the frequencies of occurrence of several intermediate steps in solutions of the problems. These statistics were compared to detect any gross abnormalities in Newell and Simon's data, compared to a larger group of subjects who solved the problem with pencil and paper without the requirement of thinking aloud.

As Table 1 illustrates, individual protocols can often be simulated in great detail, but of course there will be differences among individuals in their problem solving methods, hence in the production systems that would describe them. For purposes of psychological theory, we are often less interested in the details of a particular simulation (except as a very strong test of the theory) than we are in the structure of a program that simulates the main mechanisms revealed in a whole set of protocols. The problem of averaging over groups of subjects can also be handled formally by comparing the statistics of behavior of a program with the statistics of the human subjects as a group. In this section, we examine the processes for comparing programs in detail with individual protocols, and in Section II.A.2 we discuss the statistical approach.

Table 1 here

Individual Protocols. Newell and Simon presented several protocols in which activities of subjects reflect processes like those in GPS. An illustration is in Table 1. A segment of one subject's protocol is shown, along with a trace of a version of GPS working on the same problem. In the protocol and the GPS trace, L0 refers to the goal expression and L1 refers to the initial expression of the problem. L2, L3, and so on refer to additional expressions that are generated by the problem solver by applying operators to L1 and other previously generated expressions. The operators that are referred to in this segment are

$$R6: A \supset B \longleftrightarrow \sim A \vee B$$

$$R7: A \vee (B \cdot C) \longleftrightarrow (A \vee B) \cdot (A \vee C) \\ A \cdot (B \vee C) \longleftrightarrow (A \cdot B) \vee (A \cdot C)$$

The protocol segment in Table 1 began near the end of the first minute of work on the problem, and occupied a little more than three minutes.

In this segment, both the subject and GPS had the goal of deleting the letter R from the initial expression. Both of the problem solvers considered rule R7 as a possible means of accomplishing this. R7 cannot be applied to L1 because its connectives are wrong, so a subgoal was set to change the connective of L1. This led to use of R6, but the two occurrences of R in the transformed expression have opposite signs. Attempts were made to change one of the signs, but this returns the horseshoe to the subexpression. At this point the subject, and the specific version of GPS that produced this run, were both unable to continue on this line of work.

This protocol and GPS trace are similar to an impressive degree of detail. However, the important finding is not the fact that the subject and GPS tried to use the same rules in the same sequence. The precise sequence of rules used by GPS can be tailored fairly arbitrarily, and indeed other versions of GPS would not try to use R6 and R7 in this

Table 1

Comparison of GPS with Protocol Data

(from Newell & Simon, 1972)

GPS Trace	Subject Protocol
L0: $\neg(\neg Q \cdot P)$	
L1: $(R \supset \neg P) \cdot (\neg R \supset Q)$	
Goal 1: Transform L1 into L0	
Goal 2: Delete R from L1	
.	
.	
Goal 2: (reinstated)	Now I'm looking for a way to get rid of the horseshoe inside the two brackets that appear on the left and right sides of the equation. And I don't see it.
Goal 9: Apply R7 to L1	Yeh, if you apply R6 to both sides of the equation,
Goal 10: Change connective to V in left(L1)	From there I'm going to see if I can apply R7.
Goal 11: Apply R6 to left(L1)	(E writes L2: $(\neg R \vee \neg P) \cdot (R \vee Q)$)
Produce L4: $(\neg R \vee \neg P) \cdot (\neg R \supset Q)$	I can almost apply R7, but one R needs a tilde. So I'll have to look for another rule.
Goal 12: Apply R7 to L4	I'm going to see if I can change that R to a tilde R. As a matter of fact, I should have used R6 on only the left hand side of the equation. So use R6, but only on the left hand side.
Goal 13: Change connective to V in right(L4)	(E writes L3: $(\neg R \vee \neg P) \cdot (\neg R \supset Q)$)
Goal 14: Apply R6 to right(L4)	
Produce L5: $(\neg R \vee \neg P) \cdot (R \vee Q)$	Now I'll apply R7 as it is expressed. Both..excuse me, excuse me, it can't be done because of the horseshoe. So... now I'm looking...scanning the rules here for a second, and seeing if I can change the R to a $\neg R$ in the second equation, but I don't see any way of doing it.
Goal 15: Apply R7 to L5	(Sigh) I'm just sort of lost for a second.
Goal 16: Change sign of left(right(L4))	
Goal 17: Apply R6 to right(L5)	
Produce L6: $(\neg R \vee \neg P) \cdot (\neg R \supset Q)$	
Goal 18: Apply R7 to L6	
Goal 19: Change connective to V in right(L6)	
reject	
Goal 16: (reinstated)	
nothing more	
Goal 15: (reinstated)	
nothing more	
Goal 10: (reinstated)	
nothing more	

situation. The important finding involves the general character of the subject's performance, involving goals related to differences between the current expression and the problem goal and subgoals to make operators applicable. The protocol provides several clear illustrations of activities that are consistent with the hypothesis of a GPS-like problem-solving process.

Problem Behavior Graphs. It is important to consider whether activities like those in Table 1 are typical of problem solvers, or are relatively rare. Newell and Simon addressed this question by examining Problem Behavior Graphs (PBGs) obtained from the protocols of several subjects working on a moderately difficult problem.

Figure 1 here

An example of a PBG is shown in Figure 1. The numbers prefixed by B on the left correspond to lines of the transcribed protocol. This PBG was obtained from the protocol that includes the segment given in Table 1, which corresponds to the section of the PBG starting at B10 and ending just before B29. Information included in the cognitive states is in the rectangles; operators are shown on the lines that connect the rectangles. Information in the rectangles refers to new expressions that were written (e.g., L2 or L3, indicated in the protocol), or differences between a current expression and the goal that the subject was considering. For example, " Δg " refers to a difference in grouping of terms and " $\Delta c \text{ l\&r}$ " refers to the difference between connectives in the given expression and the goal of applying R7 (horseshoes in both the left and right sides of L1 and wedges or dots needed to apply R7).

Most of the operators refer to the rules; we mentioned R6 and R7 earlier. When a rule is applied successfully, there is an arrowhead on the line between rectangles. When a rule is shown with a line without an arrowhead, there was a goal to apply the rule but it did not succeed. Double lines indicate repetitions of attempts to apply rules.

The relation between the protocol and the PBG can be illustrated by examining the first few lines of Table 1 and the PBG starting at B10. "get L0" refers to consideration of the goal; this led to recognition of the difference in grouping between L0 and L1 (" Δg "). Then the subject attempted to apply R7; this led to identifying the differences in connectives noted in the third rectangle (" $\Delta g \text{ l\&r}$ "). Then an attempt to apply R6 was successful, resulting in line L2. The subject attempted to apply R7 a second time and noticed that there was a difference in the signs of the R terms in the two subexpressions (" ΔsR "). From time to time, the subject "backed up" to an earlier state, as when he decided that R6 should be applied only to the left side of L1. This is indicated by a vertical line from the cognitive state that the subject returned to. R6 was applied to the left subexpression of L1, giving line L3; then R7 was attempted again, but the subject noticed the horseshoe, an incorrect connective for R7. The subject returned to the goal of changing the sign of R in expression L2, but the search for an appropriate rule (indicated by R in a box) failed to produce anything helpful.

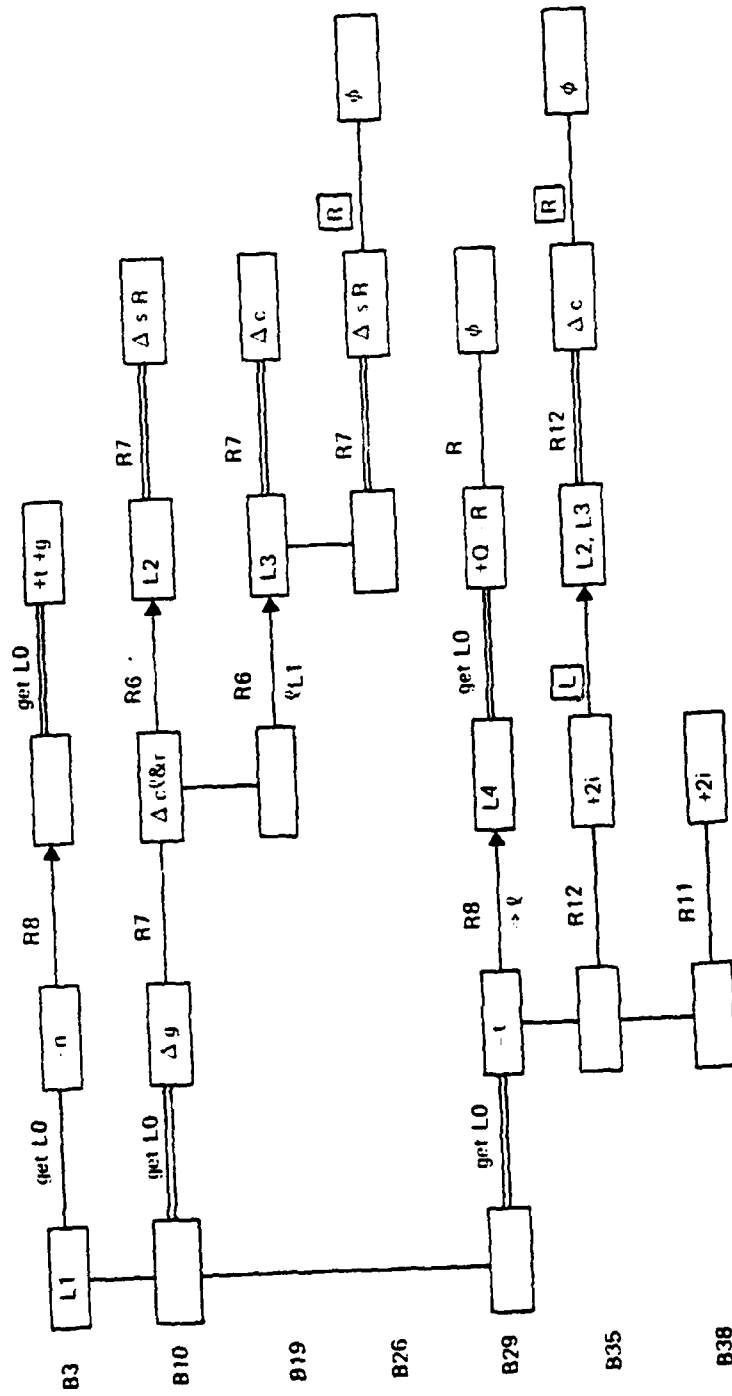


Figure 1. Problem behavior graph for a protocol, including the segment in Table 1 (from Newell and Simon, 1972).

Table 2 here

PBGs were compiled for seven subjects, working on the problem in Table 1. The transitions between states were classified, and the categories were compared with activities that occur when GPS works on a problem. The categories, and their frequencies in the seven PBGs, are shown in Table 2.

Most of the categories shown in the table correspond to GPS-like activities. Those that do not are marked with asterisks, accounting for about 18% of the transitions in the PBGs. The most interesting discrepancies involved choice of operators to avoid undesirable consequences ("avoid consequences"), and noticing features of the problem not related to the present goal ("noticing"). Simulation of these would require significant additions to GPS's problem-solving processes. The remaining discrepancies involve activities that relate to the requirement of giving protocols ("command experimenter" and "review") or where there was insufficient information in the protocol to determine whether the transition was related to one of the GPS-like categories ("other," except for those in the subcategory "noticing").

Aggregate Frequencies. The data in Table 2 were obtained from a small group of subjects who were required to think aloud as they worked. It is possible that the subjects were atypical, or that the requirement of thinking aloud caused major distortions in the way in which problem solving occurred.

Newell and Simon compared some summary statistics from their subjects with data obtained by Carpenter et al (1961) at Yale University. The number of subjects run at Yale was larger (64), and they solved the problems with pencil and paper, without thinking aloud. If the data for the Carnegie subjects did not differ from the Yale data in significant ways, then there is evidence that the general characteristics of their problem solving were not caused by individual idiosyncracies, or by the requirement of giving protocols while working on the problems.

The summary statistics involved a division of expressions into categories. Each category consists of an expression from the problem, such as the left subexpression of expression L1, and other expressions that can be formed from it by making minor transformations. Minor transformations for this purpose are those involving rules that change the order of terms, the connectives, or the signs, but do not change the terms in an expression. The data for each group of subjects are the proportions of all the expressions written that fall into the categories. The categories of expressions are listed in the left column. For example, expressions in Class L1 are those that can be formed by applying one of the minor transformations to expression L1 shown in Table 1. The categories that were used are not arbitrary; they are motivated by the observation that differences that require changing the terms in expressions are more difficult to remove, and thus require higher priority in solving the problems. (Also see the discussion of planning, which follows.)

Table 2

Total Frequencies of Occurrences of GPS-Like Mechanisms

in Seven Protocols (From Newell & Simon, 1972)

Category	Frequency	
Means-end analysis	258	
towards goal object	89	
operator applicability	151	
overcome difficulty		143
further specify		5
resolve uncertainty		3
*avoid consequences	18	
avoid difficulty		17
prepare desired result		1
Working forward	41	
systematic scan and evaluate	37	
input form similarity	3	
do something different	1	
Working backward	2	
output form similarity	2	
Repeated application	230	
after subgoal	93	
to overcome difficulty		58
to further specify		11
to resolve uncertainty		2
to avoid consequences		12
to correct error		8
to process interruption		2
implementation	97	
for plan		84
*to command experimenter		13
*review	40	
Other	27	
*noticing	6	
*repeated application	11	
*new application	10	
Total	558	

Table 3 here

Data for the problem in Table 1 are shown in Table 3. The agreement between the two groups of subjects was not exact, but the comparison does not indicate major differences in problem-solving processes. A statistical test shows that the difference between the category frequencies in the two groups was not significant ($\chi^2(4) = 8.86$; $p > .05$).^{*1}

Table 4 here

Data are shown in Table 4 for a somewhat harder problem, in which the given expression was $L1 = (P \vee Q) \cdot (Q \supset R)$, and the goal was $L0 = P \vee (Q \cdot R)$. Again, the agreement is not exact, but the difference is not large enough to reject the hypothesis that the two sets of responses were produced by a single underlying process ($\chi^2(8) = 15.27$, $p > .05$).

Planning Strategy. A second strategy of broad applicability and wide use that was identified in the logic protocols is planning. The idea underlying the planning strategy is that some gaps between the initial situation and the goal are more important and potentially harder to remove than others. If the problem space is simplified by abstracting the problem expressions, removing from them the less important features, the simplified expressions will define a much smaller space through which the search can be conducted more expeditiously. If a solution can be found to the simplified problem, then the omitted details can be restored and this solution used as a guide for searching in the original problem space.

To use the planning strategy subjects must not only be able to apply means-ends analysis, but must have enough knowledge of the problem space to be able to distinguish "important" from "unimportant" differences between expressions. For example, in the domain of logic, subjects gradually learn that it is easier to change the connectives in logic expressions than to change the letters. The planning space is then a space in which expressions like $(R \supset \sim P) \cdot (\sim R \supset Q)$ are replaced by $(RP)(RQ)$. The sequences of proof steps in the original space, $R \supset \sim P$, $\sim R \supset Q$, $\sim Q \supset R$, $\sim Q \supset \sim P$, $Q \vee \sim P$, $\sim(\sim Q \cdot P)$, becomes the simpler sequence in the planning space, RP , RQ , PQ . The second step of the search in the planning space corresponds to two separate steps in the original space, and the third step in the planning space corresponds to three steps in the original space -- a reduction of one-half in the length of the derivation, and of a much larger factor in the amount of search required to find it.

Evidence for planning was obtained in protocols like the following, obtained in a problem with four given expressions: $L1 = P \vee Q$; $L2 = \sim R \supset \sim Q$; $L3 = S$; $L4 = R \supset \sim S$; and the goal: $L0 = P \vee T$. Rule R9, mentioned in the protocol, is $A \rightarrow A \vee X$, a rule for adding a term to an

^{*1.} The independence assumption of the chi-square test was not met in these data, since several expressions were written by each subject. However, this would generally make it more likely that a significant difference would be obtained, so the conclusion seems warranted.

Table 3
Proportions of Expressions

Class of Expressions	Carnegie (78 expressions)	Yale (519 expressions)
L1	.37	.29
left of L1	.14	.16
right of L1	.12	.22
L0	.24	.17
other	.13	.16

Table 4
Proportions of Expressions

Class of Expressions	Carnegie (97 expressions)	Yale (487 expressions)
L1	.33	.28
extended L1	.02	.04
left of L1	.14	.19
right of L1	.14	.15
(RVP)	.13	.07
(PVQ) • (PVR)	.03	.01
L0	.03	.01
Rule 9	.16	.18
other	.01	.07

expression.

Well, one possibility right off the bat is when you have just a PVT like that the last thing you might use is that R9. I can get everything down to a P and just add a VT. So that's the one thing to keep in mind.

Well, maybe right off the bat, I'm kinda jumping into it, I maybe can work everything down to just a P; I dunno if that's possible. But I think it is, because I see that steps 2 and 4 are somewhat similar; if I can cancel out the R's, that would leave me with just an S and Q;

and if I have just an S and Q, I can eventually get step 3, get the S's to cancel out and end up with just a Q;

and if I end up with just a Q, maybe the Q's will cancel out; so you see, all the way down the line. I dunno, it looks too good to be true, but I think I see it already.

II.A.2. Water-Jar Problems. We now discuss an analysis of problem solving in another task. Water-jar problems, studied extensively by Luchins (1942), are transformation problems with definite goals, involving a set of three jars of different capacities. In the form studied by Atwood and Polson (1976), the largest jar is full in the initial state, and the goal is to have that water divided equally between two jars. For example, the capacities may be Jar A: 8 oz.; Jar B: 5 oz.; Jar C: 3 oz. Then in the initial state, Jar A contains 8 oz. of water, and Jars B and C are empty. The goal is to have 4 oz. of water each in Jars A and B. The problem-solving operators involve pouring water from a source jar into a target jar. Water can be poured into the target jar until it is full, if there is enough water in the source jar; water can be poured out of the source jar until it is empty, if there is enough room in the target jar. Intermediate actions are not possible.

In the water-jar task, differences between any state and the problem goal consist of discrepancies between the contents of the three jars in that state and the contents that are specified in the goal. Atwood and Polson hypothesized that subjects would judge their progress by combining the discrepancies, forming an overall evaluation function for the current state, and would try to select moves that would improve the value of this function. They assumed that the evaluation of a specific state i was

$$e_i = |C_i(A) - G(A)| + |C_i(B) - G(B)|,$$

where $C_i(A)$ and $C_i(B)$ are the actual contents of Jar A and Jar B in state i , and $G(A)$ and $G(B)$ are the contents of Jar A and Jar B in the goal state. (The contents of Jar C are redundant with those of A and B.)

Atwood and Polson formulated a process model, based on the means-ends strategy of attempting to reduce the evaluation to zero. They assumed that at each move subjects consider various pouring operations that could be made legally, and try to choose one that will make the evaluation function smaller, or at least not increase its current value by more than a threshold amount.*2 Atwood and Polson also made specific assumptions about memory capacity; they assumed a limited short-term memory for holding information about states that would be produced by alternative moves, and they assumed that each state reached in solving the problem was stored in long-term memory with a fixed probability.

The model also specifies a sequence of processes for selecting a move. The sequence includes calculating the evaluation function for alternative moves, storing information about alternatives in STM, recognizing states that have occurred before on the basis of information in LTM, and deciding whether to make a given move under consideration. The assumptions of the model allow for several possibilities. A move might be selected if it leads to an acceptable state; this was assumed to be less likely if the state was recognized as having occurred before. The moves stored in STM may be examined, with selection of the stored in LTM from previous occurrences, a move may just be chosen at random from the set of possible moves, or the subject may decide to return to the initial state of the problem.

Atwood and Polson tested their model with data obtained from groups of human subjects who solved different versions of the problem. Problems were presented at computer terminals and records were kept of the moves made by each subject. The model was implemented as a computer program which was run with various values of the parameters. Because the model contains probabilistic processes, it does not produce a single sequence of moves in solving a problem. The model was run many times with each set of parameter values, and a summary of its performance was obtained, consisting of the average frequency of each of the possible problem states. A set of parameter values was chosen for which the set of frequencies for two problems (jar sizes of 8,5,3 and 24,21,3) approximated the frequencies obtained from the human subjects. The parameter values that were chosen seem quite reasonable. The size of STM was set at three alternative moves; states reached in the problem were stored in LTM with probability .90; and the threshold of acceptability for a new state was set at 1.0 above the value of the current state.

Figure 2 here

Results of the simulation are shown in Figure 2. Each set of predictions was based on running the model 250 times. The data for each problem were from a group of about 40 subjects, different from the data used to estimate the parameters. One problem, (8,5,3), was used in estimation, but the other three problems were different. The model correctly predicted the order of difficulty of these four problems. For two of the problems, (8,5,3) and (12,7,4), the detailed predictions of response frequency were satisfactorily close to the data by a statistical test. For the two harder problems, although the general shapes of the frequency distributions agreed with the data, the model erred in predicting

*2. This strategy differs from the means-ends strategy of GPS in one significant respect. GPS considers all the ways in which the current state and the goal differ, and selects a move to reduce the most important of these qualitative differences. Atwood and Polson's model combines the differences into a single numerical index, the value of the evaluation function, and tries to reduce that difference by at least a threshold amount. This difference probably does not have a significant effect on predictions of performance in the water-jar task, but there are situations in which strategies based on global evaluations and on individual qualitative differences would lead to significantly different performance.

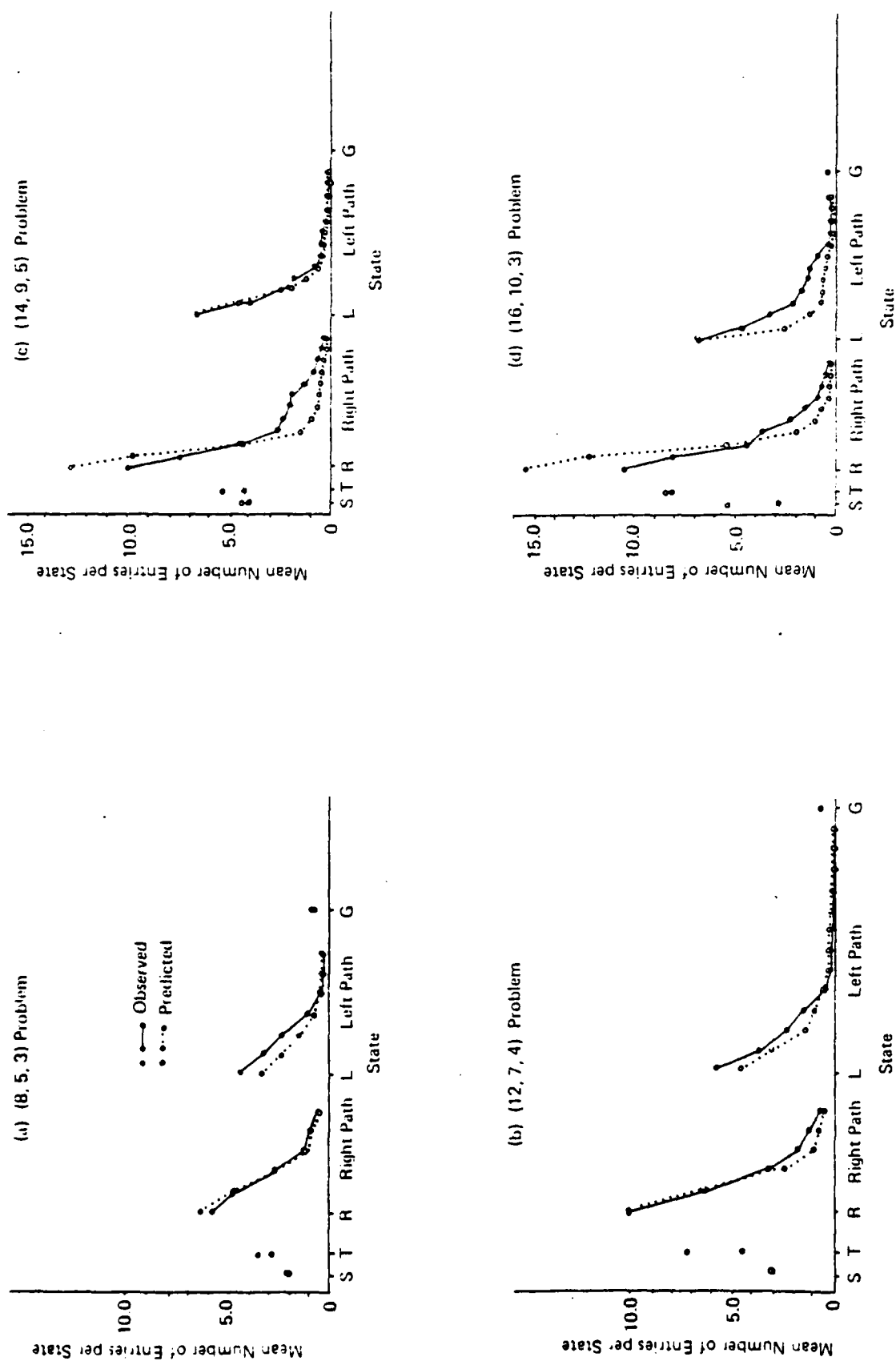


Figure 2. Observed and predicted values of mean visits per state for four water-jar problems (from Atwood & Polson, 1976).

too many returns to states at the beginning of a path that led to the goal. As Atwood and Polson noted, this defect could be corrected by making the probability of recognizing a previous state depend on the number of times it has been encountered.

Conclusions. Section II.A has been concerned with problem solving in situations that are novel to the problem solver, in which a definite goal and the set of legal problem-solving operators are described by the instructions. The situation requires using some general problem-solving strategy. The findings show that in situations of this kind, the strategy of means-ends analysis represents the major feature of human problem-solving performance. In this section we have discussed evidence consisting of individual thinking-aloud protocols and aggregate response frequencies in two tasks. Findings fitting this general pattern have been obtained in a wide range of problem-solving tasks, including puzzles such as the Tower of Hanoi (Anzai & Simon, 1979) and physics textbook problems (Simon & Simon, 1978), which we discuss below in Section II.C.

Means-ends analysis is perhaps the single most important strategy that people employ for searching selectively through large problem spaces. The selectivity is powerful because it points search in the direction of the goal, selecting operators on the basis of their relevance to reducing the distance from that goal. Use of means-ends analysis requires some domain-specific knowledge; for example, it can be employed efficiently only if the subject has learned enough about the problem domain to have associated particular differences with particular operators for removing them. However, it is basically a "weak method," applicable in situations where the problem solver has little specific knowledge based on experience in the problem domain.

II.B. Domain-Specific Knowledge for Familiar Problems with Specified Goals

We now turn to problems that are solved by individuals who have specialized knowledge, acquired either through instruction or practice. We will discuss problem solving in a domain of school mathematics, high school geometry. Then we will discuss a phenomenon that has been salient in the problem-solving literature, problem-solving set or Einstellung, which we interpret as resulting from domain-specific knowledge structures.

II.B.1. Geometry Exercises. In school subjects such as geometry, knowledge for solving problems is imparted intentionally through instruction. Research conducted by Greeno (1978) had the goal of investigating and characterizing the knowledge that is acquired by students who learn successfully in the course.

The main data were obtained in a series of interviews conducted approximately once each week with six students who were taking a standard high school course in geometry. In each interview, an individual student worked for about 20 minutes, during which he or she typically solved three or four problems. Most of the problems that were solved were typical of homework or test problems that the students were working on at that time in the course. Students were asked to think aloud as they worked, and their protocols were recorded and transcribed.

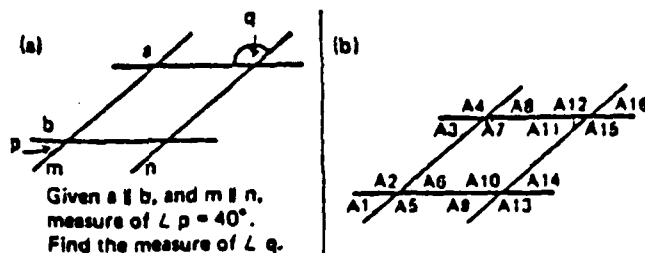
Figure 3 here

One of the problems solved in an early session (during the second month of the course) is shown in Figure 3. The problem as it was presented is shown in the upper left. The upper right diagram provides notation for referring to the various angles in the diagram. The seven steps shown below the diagrams are a formal solution with inferences and justifying reasons. The students were not required to write the solution steps of this problem formally but they were required to state aloud the intermediate inferences that they made. Most of the students solved the problem in Figure 3 correctly. We will discuss specific aspects of their solutions below. They were generally similar to the solution shown in Figure 3.

The solution shown in Figure 3 was given by a computational model called Perdix that was formulated to simulate the students' performance. The structures and processes represented in Perdix are hypotheses about the knowledge that students acquire in a geometry course.

Problem-Solving Knowledge. Perdix contains three kinds of knowledge, all represented as production rules: (1) problem-solving operators that make inferences, (2) perceptual concepts that recognize patterns in diagrams, and (3) strategic processes that set goals and select plans for problem-solving activity.

Problem-solving operators in geometry correspond to the theorems, postulates, and definitions that are used as reasons to justify steps in a problem solution. Examples include "Vertical angles are congruent" (a theorem), "Corresponding angles are congruent" (a postulate), and "If two angles are supplementary, the sum of their measures is 180° " (a



Statement	Reason
1. meas. $A1 (P) = 40^\circ$	1. Given
2. $A1 (P) \cong A6$	2. Vertical angles
3. $A6 \cong A8$	3. Corresponding angles
4. $A8$ supplem. $A12 (Q)$	4. Interior angles on same side
5. $A6$ supplem. $A12 (Q)$	5. Substitution
6. $A1 (P)$ supplem. $A12 (Q)$	6. Substitution
7. meas. $A12 (Q) = 140^\circ$	7. Definition of supplem.

Figure 3. A solved problem in geometry. ($A1$, $A6$, etc. in the solution refer to the positions of angles in the upper right diagram.)

definition). When the antecedent of one of these propositions is satisfied in a problem, then the consequent can be inferred. For example, because A1 and A6 are vertical angles in Figure 3, the inference that A1 and A6 are congruent is permitted. The propositions that correspond to the problem-solving operators are prominent in geometry instruction. They are represented in Perdix as production rules, with the antecedents as conditions and the relations that can be inferred as actions.

Patterns of information in the problem have to be recognized to determine that a problem-solving operator can be applied. For example, to apply the inference rule "Vertical angles are congruent" in Figure 3 and thus infer that A1 and A6 are congruent, the problem solver must first recognize that A1 and A6 are vertical angles. In the geometry course, perceptual concepts are taught with examples using diagrams. In Perdix, knowledge for recognizing patterns is represented by discrimination networks, similar to the structures in the Elementary Perceiver and Memorizer, EPAM (Feigenbaum, 1963) and the Concept Learning System, CLS (Hunt, Marin & Stone, 1966). Perdix's recognition system is based on features of a diagram, such as sides of two angles that are collinear, along with other information that may be given or inferred, such as statements that lines are parallel or perpendicular. An example is shown in Figure 4, which represents the process that can recognize a pair of vertical angles, a pair of angles formed by bisecting an angle, and other patterns that involve pairs of angles that have a single vertex.

Figure 4 here

Strategic knowledge is needed for setting goals that organize problem-solving activity. In the example problem of Figure 3, the main goal is to find the measure of angle Q. This cannot be achieved directly, and the problem solver must know that a way of finding the measure of an angle is to find a quantitative relationship (e.g., congruent or supplementary) of the unknown angle with one that has a known measure. This can be represented as a production: when the current goal is to find the measure of an angle, and the measure of another angle is known, set a subgoal of finding a quantitative relation between the unknown angle and the known angle.

Table 5 here

The importance of strategic knowledge is illustrated in the protocol in Table 5. The student was working on the problem shown in Figure 3. The student marked several angles in a copy of the diagram; these are indicated in the protocol in parentheses in relation to the diagram in the upper right part of Figure 3. For example, "P would equal one (\rightarrow A1)" indicates that a label "1" was written on the angle in the student's diagram at position A1.

The student seems to have known the problem-solving operators and the geometric patterns needed to apply them (this was confirmed in another part of the interview) but was unable to solve the problem. The most likely hypothesis is that the student lacked knowledge of the problem-solving strategy needed in this problem. The strategy involves forming a chain of

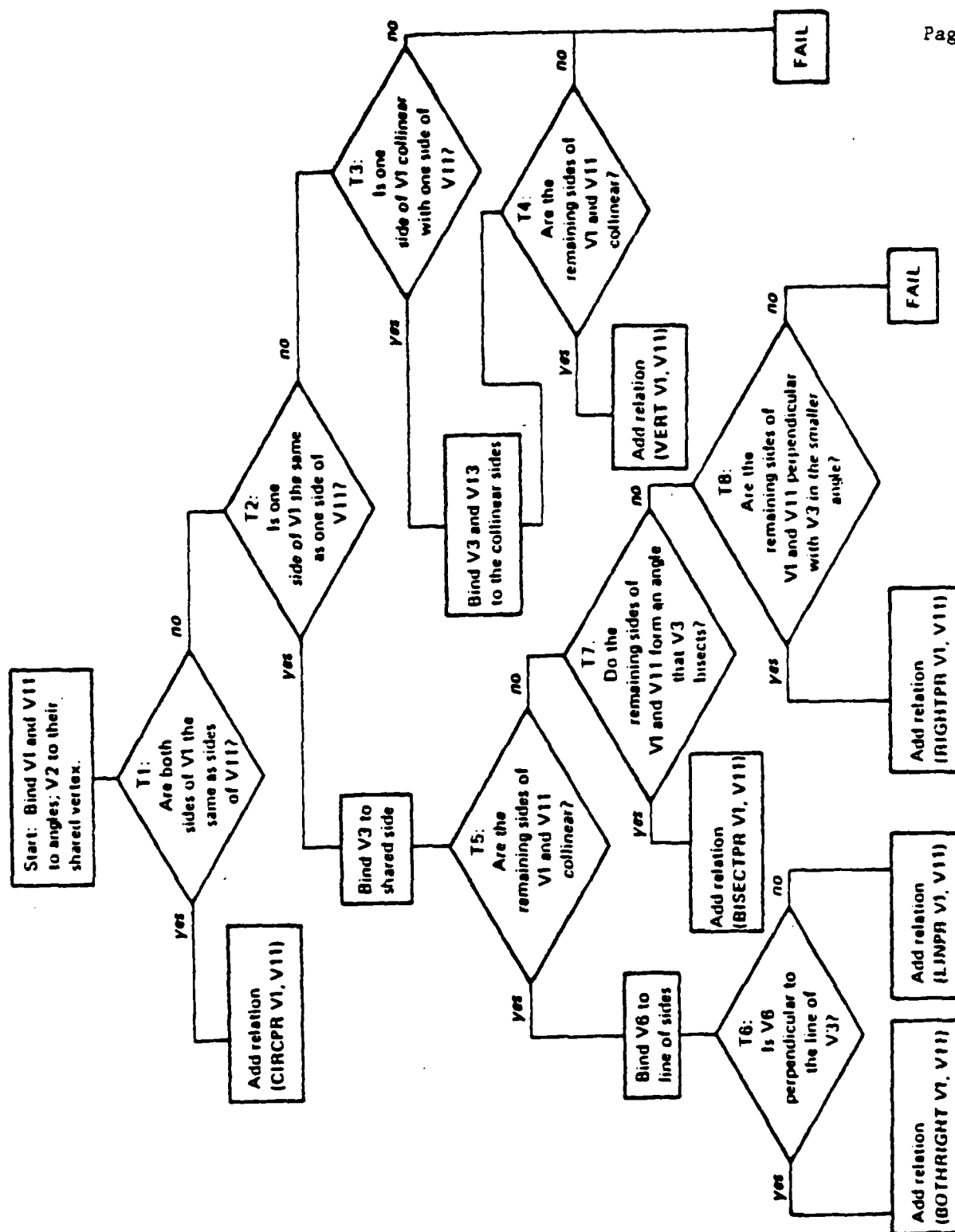


Figure 4. Pattern-recognizing system in Perdix for analyzing angles with a shared vertex. (from Greeno, 1978).

Table 5

Protocol of an Attempt to Solve Figure 3

- S: All right. I would put, like, P would equal one ($\rightarrow A1$).
- E: Okay.
- S: And then, two ($\rightarrow A6$).
- E: Put in two there, right.
- S: And then three ($\rightarrow A15$); no, wait -- three ($\rightarrow A15$) and four ($\rightarrow A12$), I guess.
- E: Okay. Now, why did you put two there?
- S: Well, I don't know. It could have something to do with vertical angles.
- E: Okay.
- ...
- S: All right, the first thing I guess I should try to do, I would try to find if there were any alternate interior or corresponding angles?
- E: Okay.
- S: Or any of those.
- E: Mm-hm.
- S: I guess I would say that ... well, wait a minute. I guess maybe I would put five there ($\rightarrow A16$).
- E: Okay.
- S: I don't know if I would need this.
- E: Okay.
- S: These two are supplementary.
- E: Right.
- S: That doesn't help much. And then, the measure of angle five ... would it equal the measure of angle one?
- E: Well, you might have to work that out.
- S: How ... if this equals ... this equals forty.
- E: That's right.
- S: Oh, all right. Wait, the measure ... I can't, I don't know. I don't know how to do these.
- E: Okay.

angles that are related by congruence. Knowledge of this strategy involves setting a series of goals; when the problem requires a relation between two angles, and none can be recognized, then find an angle related to one of them by congruence and try to relate that angle to the other angle. This strategic procedure can be applied recursively until an angle is found that is related to the goal angle by one of the geometric relations from which a quantitative relation can be inferred.

Four of the six students who were interviewed in Greeno's study solved the problem in Figure 3 successfully, apparently having acquired the strategy of forming a chain of congruent angles. About a week after giving the protocol in Table 5, that student also was successful in solving a problem that required the chaining strategy. The students differed in the specific sequences of angles that they used, which could be the result of differences in the way that they scanned the diagram looking for angles to add to the chain, or to differences in the ease with which different students recognized various geometric patterns. About a week after giving the protocol in Table 5, that student was successful in solving a different problem that also required the chaining strategy.

In geometry instruction, very little strategic knowledge is taught explicitly; it has to be inferred by the students from example problems. We believe that this is a common feature of instruction in domains requiring acquisition of knowledge for problem solving, and we consider the explicit teaching of problem-solving strategies as a potentially productive development for instruction, based on the results of basic research on cognitive processes in problem solving.

Strategic knowledge is represented in Perdex by productions that select plans for work on problems. A plan is a general approach to the problem, based on information in the problem situation. GPS forms such plans using its general planning strategy, described on pages 19-20. Perdex has specific cognitive structures for plans that are used frequently for geometry problems. Forming a chain of congruent angles is one such plan. Another is using congruent triangles to prove that two angles or two line segments are congruent.

The organization of planning knowledge in Perdex is similar to that developed by Sacerdoti (1977), called a procedural network. In a procedural network, there are units of knowledge corresponding to actions at different levels. Each of these knowledge units includes information about the prerequisites and consequences of an action that can be performed. In Perdex, knowledge of each plan includes information about goals that can be achieved using the plan (its consequences), conditions in problems that make the plan promising (its prerequisites), and subgoals that should be set if the plan is adopted.

Perdex's strategic knowledge constitutes the main way in which it differs from GPS. Strategic knowledge in GPS is the general means-ends strategy that can be used in any domain for which the problem solver is taught the operators, together with the productions that connect operators with differences, and given the goal of a problem. The hypothesis represented in Perdex is that instruction in a domain such as geometry leads to acquisition of strategic knowledge specific to the domain, such as the schematic knowledge that represents plans to use chains of congruent

angles or congruent triangles. Both GPS and Perdix construct plans that are more general than the actions that must be performed in solving the problem. The difference is that GPS forms plans using its general means-ends strategy, while Perdix's plans are based on knowledge of specific geometry strategies.

When GPS plans, it use the strategic process of means-ends analysis in a problem space that contains features taken directly from the basic representation of the problem. GPS's planning space can be acquired by learning which features of objects should be given first priority. In Perdix, planning uses schematic knowledge of specific methods applicable to problems in the domain of geometry. These schemata include general subgoals, such as proving that triangles are congruent or finding an angle with a relation based on parallel sides, that can be used as intermediate steps. The associations of these subgoals with the goals that they help to achieve have to be acquired by students; they are not explicitly given as goals of problems in which they are used.

Solution of Ill-Structured Problems. A hypothesis that is consistent with the analysis of geometry problem solving is that domain-specific strategic knowledge may provide the main basis for solving ill-structured problems. Problems may lack definite structure for many reasons. One important source of indefinite structure is that a problem may require knowledge from several different sources, so its solution requires coordinated work in several disparate problem spaces (Simon, 1973).

A modest form of this kind of problem arises in geometry, involving problems that require construction of auxiliary lines. The problem space that is presented, including a diagram, given information, and a goal to be proved, must be augmented in order for the problem to be solved. Greeno, Magone, and Chaiklin (1979) proposed that solution of such problems can be based on an individual's knowledge of plan schemata. In the model Perdix, the need for an auxiliary line is recognized when a plan's prerequisites are partially satisfied in the problem situation. This leads to definition of a subproblem; the goal is to complete the pattern of features that constitute the prerequisites, and this goal is achieved in a problem space with operators that are appropriate for that goal.

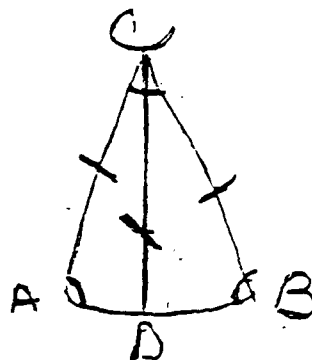
Figure 5 and Table 6 here

An example is shown in Figure 5, the drawing and written work of a student on the following problem: "Prove that if two sides of a triangle are congruent, then the angles opposite those sides are congruent." The protocol given by this student is in Table 6. After drawing the triangle ABC, the student added the line CD, which is not specified in the initial problem space. The student's comments at *1 and *2, along with the retrospective comment at *3, provide evidence that construction of the auxiliary line was related to a plan of proof involving congruent triangles, and the construction completed a pattern that is required for that plan to be applied, that is, the presence of two triangles in the diagram. Perdix simulates solutions like this with a process of patterns recognition that identifies partial patterns of two triangles missing a line, and uses special problem-solving operators to complete the patterns.

If 2 sides of $\triangle \cong$, then the angles opposite those sides are \cong

Given $\triangle ABC$; $\overline{AC} \cong \overline{BC}$

Prove: $\angle A \cong \angle B$



1. Given

1. $\triangle ABC$; $\overline{AC} \cong \overline{BC}$
2.

Figure 5. Drawing and written work on the problem, "Prove that if two sides of a triangle are congruent then the angles opposite those sides are congruent."

Table 6

Protocol for the Problem of Figure 5

S: Okay, if two sides of a triangle are congruent, so . . . draw a triangle.

E: Okay.

S: Then the angles opposite those sides are congruent. Okay, so, like, if I have . . . given: triangle ABC--I'll letter it ABC.

E: Right.

S: And then I have . . . prove: . . . do I already have these two sides given? Okay. Two sides of a triangle are given.

E: Mmm-hmm.

S: Let me go back to my given and say that segment AC is congruent to segment BC.

E: Okay.

S: And I want to prove that angle A is congruent to angle B.

E: Good.

S: All right. Let me write down my given. Okay. And mark my congruent sides. Okay, so, I want to prove that angle A is congruent to angle B. Now, let's see. Do you want . . . ?

E: Yeah. Why are you drawing a line there?

*1 S: I don't know yet.

E: Oh, that's okay. Don't erase it.

S: I'm going to do it, no, I just . . .

E: Oh, okay, fine.

S: Okay . . . okay, then I could . . . if I drew a line . . .

E: Mmm-hmm.

*2 S: That would be the bisector of angle ACB, and that would give me . . . those congruent angles . . . no. (Pause.) Yeah, well, that would give me those congruent angles, but I could have the reflexive property, so this would be equal to that. Okay, I've got it.

E: Okay.

S: Okay.

E: Now, before you go ahead and write it all down, when you said you were going to draw the line . . .

S: Yeah.

E: And I said why are you doing that, and you said you didn't know yet, what do you think happened to give you the idea of making it the bisector?

*3 S: Okay, well, I have to try to get this . . . I have to try to get triangle ACD congruent to BCD. Because, if I do that, then angle A is congruent to angle B because corresponding parts of congruent triangles are congruent.

E: So you were drawing the line to give yourself triangles, is that the idea?

*4 S: No, to . . . to get a side that was in both triangles.

E: Okay.

S: And to get congruent angles.

E: So that's why you drew it as the bisecotr.

S: Yeah.

Another way in which problems can be ill-structured involves the way in which goals are formulated. Goals in well-structured problems are presented as specific objects (e.g., a specific logic expression to be derived or a specific distribution of water among some jars). In ill-structured problems, goals are often underdetermined, with several alternative ways in which they might be satisfied. Examples are frequently cited from art or science, such as the goal to compose a fugue, or to design an interesting experiment. In school geometry, the goals of problems are usually well specified, but a subgoal that arises in many problems functions as an indefinite goal for experienced problem solvers. This is the goal of proving that two triangles are congruent. There are several ways in which congruence of triangles can be proved, involving different patterns of congruent components such as side-side-side, side-angle-side, and so on. Beginning learners treat these as definite subgoals, trying one after another until one is found that works (Anderson, Greeno, Kline & Neves, 1981). However, more experienced students do not mention specific patterns in their protocols, and appear to engage in relatively diffuse search for congruent components of triangles with a kind of monitor that identifies whatever pattern of congruent components happens to emerge. Greeno (1976) hypothesized that experienced students acquire an integrated structure of knowledge in the form of a pattern-recognizing system that represents the goal of proving that triangles are congruent. A version of this that was implemented in Perdex is shown in Figure 6.

Figure 6 here

Acquisition of Problem-Solving Skill. An important question is how the knowledge required for solving problems in a domain such as geometry is acquired. We discuss studies of learning involving the three kinds of knowledge for problem solving: problem-solving operators, perceptual concepts for pattern recognition, and strategic knowledge.

Processes of acquiring problem-solving operators were analyzed by Anderson (1982), based on observations of three students as they studied and worked problems in the early sections of a geometry text. Anderson simulated processes of acquiring problem-solving skill in a version of his ACT model (cf. Anderson, 1983).

A major aspect of Anderson's model is a process that acquires cognitive procedures from declarative information. ACT learns new procedures by working on problems. When ACT encounters a problem for which it has not learned a procedure, it uses general problem-solving methods along with information that is available as it is in a text. For example, a geometry problem may require finding a theorem that can justify a step in a proof. ACT has a general procedure for searching in a list of theorems and matching features of theorems to the information in a problem. When an applicable theorem is found, ACT asserts that theorem to solve that part of the problem.

ACT has a learning process called proceduralization, which forms new production rules that are added to ACT's procedural knowledge. A new production can be formed when a theorem has been found and applied successfully in problem solving. The new production has conditions corresponding to selected features in the problem situation, and an action

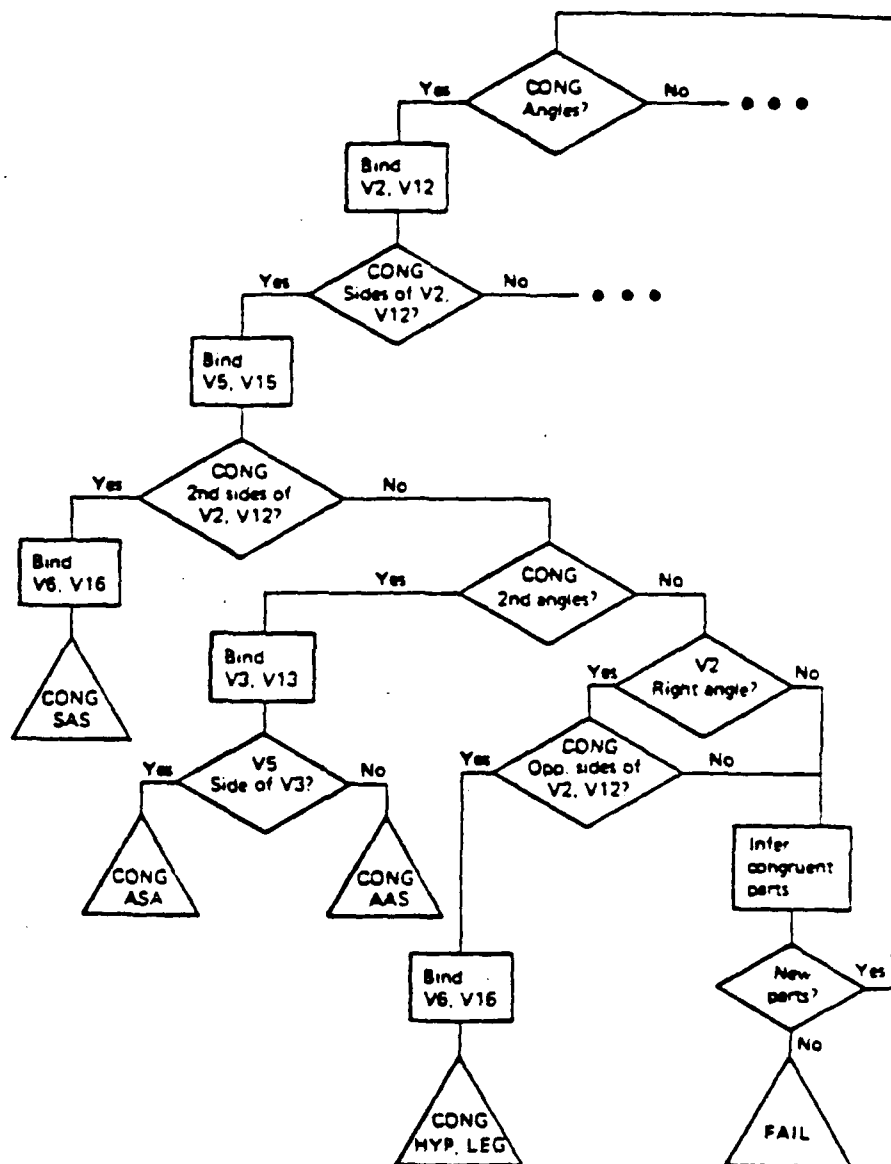


Figure 6. Part of Perdix's goal structure for proving congruence of triangles, represented as a pattern-recognizing system. (Abbreviations: Cong-congruent; SAS=side-angle-side; ASA=angle-side-angle; AAS=angle-angle-side; HYP; LEG=hypotenuse-leg.) (from Greeno, 1976).

of asserting the theorem. The production is a new problem-solving operator; ACT has acquired a new ability to assert a theorem in appropriate conditions, without having to search in the list of theorems in the text. It has learned the theorem, not in the sense of having memorized it, but in the sense of being able to recognize when it is applicable and to apply it.

Acquisition of perceptual concepts for pattern recognition in problem solving was studied by Simon and Gilmarin (1973) in the domain of chess. The learning mechanism used was adapted from the EPAM model (Feigenbaum, 1963), which simulates acquisition of discrimination networks like Figure 4. Simon and Gilmarin developed an EPAM-type model that acquired knowledge of patterns of chess pieces from presentations of board positions. This knowledge was used to simulate performance in a task of reconstructing positions after brief presentations, a task known to differentiate players according to their level of skill (deGroot, 1965; Chase & Simon, 1973; also see Section III.B.2).

Acquisition of strategic knowledge for solving problems has been studied empirically by Schoenfeld (1979). Four students in upper-division college mathematics courses were given special instruction in the use of five heuristic strategies for working on problems: drawing a diagram, arguing by induction, arguing by contradiction or contrapositive, considering a simpler problem with fewer variables, and establishing subgoals. Each strategy was presented in a training session, lasting about one hour, including explanation of conditions in which the strategy is useful as well as practice in using the strategy. Students took a pretest and a posttest with problems not included in the training. Students who received the special training had a list of the strategies available during the posttest and were reminded from time to time to try to use one of the strategies if they were not progressing well on a problem. Performance of these students was superior to that of another group of students who had worked on the same training problems as the instructed group, but without explanation of the strategies. Thinking-aloud protocols confirmed that students considered and used strategies that they had been trained to use. The training was especially effective with strategies that have clear cues for their application: the fewer-variables strategy, cued by the presence of many variables, and arguing by induction, cued by an integer argument.

Processes of acquiring strategic knowledge have been addressed in theoretical analyses by Anzai and Simon (1979) and by Anderson, Farrell, and Sauers (1982). Anzai and Simon observed and simulated acquisition of a strategic concept in the Tower of Hanoi puzzle. The concept involves movement of a set of disks requiring a sequence of individual moves, with the sequence considered as a global action. Anderson et al. simulated acquisition of knowledge for applying techniques in learning to program in LISP. In both of these theoretical analyses, important factors in acquiring strategic knowledge are activation of a problem goal that can be achieved by a sequence of actions and acquisition of a production in which the action of setting that goal is associated with appropriate conditions in the problem situation.

II.B.2. Einstellung (Set) The context in which problem solving occurs may have an important influence on the process. As a consequence of previous tasks in which a subject has been engaged or previous stimuli that have been presented, certain responses may become more readily and speedily available and others less readily available. The subject has acquired a "set" for the familiar stimuli and responses.

One experimental design that has been used often to demonstrate the effects of set is to present subjects with a sequence of tasks that induce set, then a new sequence of tasks in which the set that has been induced either facilitates or impedes performance in comparison with control subjects who were not exposed to the first sequence. Luchins (1942) conducted a well-known set of experiments using this design, with water-jar tasks.

In Luchins' version of the water-jar task, subjects must measure out a specified amount of water, using a given set of ungraduated measuring jars. A source of water is assumed to be available, so that any of the jars can be filled to its capacity if the subject chooses to do that. In addition, water can be poured from one jug to another, until the target jar is filled or the source jar is empty, and the contents of a jar can be discarded.

Table 7 here

The series of problems that Luchins used is in Table 7. Here, all the problems except the first and the ninth can be solved by filling jar B, then pouring from it to fill A, and then filling C twice ($X = B - A - 2C$). But Problem 5 and Problems 7 through 11 can also be solved using only jars A and C -- by either adding the contents of C to the contents of A, or subtracting the contents of C from A, and for Problem 9, the $B - A - 2C$ procedure does not work.

Subjects given Problems 7 through 11 immediately after solving Problem 1 generally use the two-jar procedure just described. Subjects who are first given Problems 1 through 6 generally use the $B - A - 2C$ procedure, which is more complex than necessary for Problems 7 through 11, and they have considerable difficulty with Problem 9.

Set effects can be the result of several cognitive processes; we will discuss three that have been put forward.

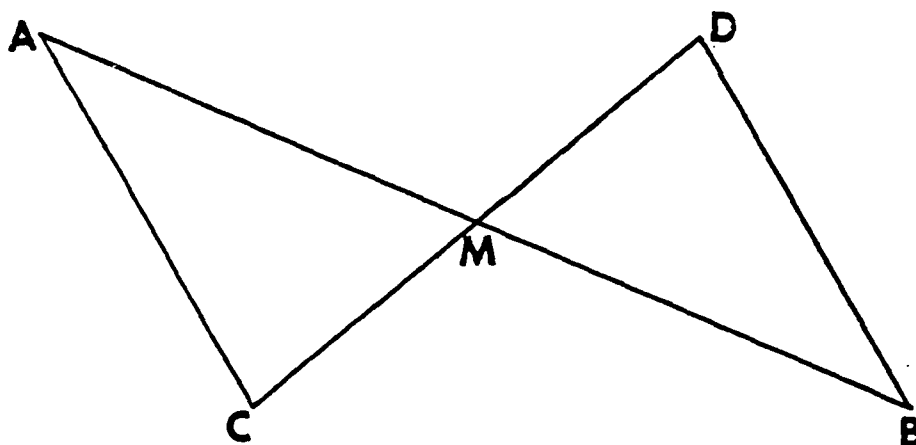
First, set may be the result of a bias in retrieving knowledge structures from memory. A standard assumption is that the alternative concepts or cognitive procedures that might be retrieved have varying strengths or levels of activation which determine the probabilities of their retrieval. If a cognitive unit has been used successfully several times in the immediate past, this results in a relatively high level of activation for that unit.

Figure 7 here

Table 7

Problems Used by Luchins (1942)

Problem Number	Measuring Jugs			Required Amount
	A	B	C	
1	29	3		20
2	21	127	3	100
3	14	163	25	99
4	18	43	10	5
5	9	42	6	21
6	20	59	4	31
7	23	49	3	20
8	15	39	3	18
9	28	76	3	25
10	18	48	4	22
11	14	36	8	6



Given: M is the midpoint of \overline{AB} and \overline{CD} ;

$\overline{AC} \cong \overline{BD}$.

Prove: $\angle AMC \cong \angle BMD$

Figure 7. An Einstellung problem in geometry.

Schemata used in planning provide one kind of structure that can account for set. An example is in the domain of geometry, where Greeno, et al. (1979) developed a simulation model with planning schemata, described in Section II.B.1. Luchins (1942) included a study of geometry problem solving in his investigations of Einstellung. Figure 7 shows the kind of problem used as a test. The proof can be obtained in one step; $\angle AMC$ and $\angle BMD$ are vertical angles. However, if subjects were first given a series of problems where they used congruent triangles in proofs, they were likely to construct the more complex proof for Figure 7 in which triangles AMC and BMD are proved congruent by Side-Side-Side. An explanation is provided if we assume that students have a schema corresponding to the plan of using congruent triangles for a proof, and that this schema has a high level of activation because of its use in the initial series of problems. Greeno et al. (1979) reported an experiment with a test problem that could be solved either using congruent triangles or angles formed by parallel lines, and either method required construction of an auxiliary line. Subjects were given series of problems before the test problem involving either congruent triangles or parallel lines, and were strongly biased toward solutions of the same type they had been giving.

Set based on activation may either facilitate task performance or impede it, depending on whether the memory elements that are activated contain the information that is needed for performance. Sweller and Gee (1978) showed that the tendency to use a previously successful rule can greatly facilitate solution of a relatively complex problem, presumably by eliminating the need to search in a large space of possibilities, even though in the same situation it prevents subjects from noticing a simpler solution method. Such situations are common, since set is bound to arise wherever memory organization is not neutral with respect to the problem-solving process -- that is, wherever there are alternative ways of storing information in memory, one of which may be more conducive to retrieval in a given problem context than another.

A second possible explanation of Einstellung is provided by composition of productions, investigated first by Lewis (1978). Composition is a process in which a newly acquired production performs actions that required two or more productions in the previous knowledge structure. Composition generally makes performance more efficient by providing a way to act directly rather than requiring several steps to achieve a goal. The new productions created by composition usually have conditions that are relatively specific, and in some production systems (including ACT) this leads to their being preferred to productions with less specific conditions. Anderson (1982) noted that this would simulate the performance observed by Luchins (1942) on problems like Figure 7.

Third, some set-like phenomena could also be produced by the basic problem-solving procedure that a subject uses. We have already noted that subjects very frequently use the heuristic of means-ends analysis -- that is, comparing situation with goal and taking an action that seems to reduce the difference between them. In their analysis of behavior of subjects solving water-jar problems, Atwood and Polson (1976) showed that where alternative actions could be taken, most subjects selected the one that led to a situation that was most like the goal situation. Like the more specific sets induced by Luchins' manipulation, this general set to pick paths that lead toward the desired goal can sometimes interfere with

problem solution. Where memory limitations prevent subjects from looking far ahead, this goal-oriented strategy may sometimes produce a myopic preoccupation with immediate progress, and an avoidance of paths that lead to the goal only indirectly. Jeffries, Polson, Razran, and Atwood (1977) showed that, without look-ahead, subjects solving the Missionaries and Cannibals puzzle would have difficulty (as, in fact, they do) on the step where they were required to bring two persons back from the farther bank of the river to which they were trying ultimately to transport all of them.

II.C. Problems with Specified Procedures

In the tasks discussed in II.A and II.B, a definite goal is presented to the problem solver. In this section we discuss tasks in which the problem presents material for a procedure, and the task is to apply the procedure to find the result that is obtained. While the tasks discussed in II.A and II.B specify a goal and require a method to get there, the tasks we discuss now specify a method and ask where the method leads.

The tasks that we discuss come from arithmetic. Many tasks of applying procedures occur in mathematics, for example, finding a derivative in calculus or finding the product of two expressions in algebra. Some people would object that such tasks do not involve problem solving, since they require knowledge of a procedure rather than search in a space of possible solutions. On the other hand, these tasks are considered as problems by students who receive them as homework assignments (and presumably by teachers who assign them).

More significantly, the knowledge required for these procedure-based tasks is similar to the knowledge that individuals acquire when they learn to solve problems that do not specify solution methods, such as geometry proof exercises or water jar problems. Knowledge for planning in geometry constitutes a set of procedures that the student has acquired for solving various kinds of problems. Use of these procedures requires recognition of their applicability, which is not required if the problem says "subtract" or "differentiate;" however, characteristics of the procedural knowledge that have been identified by theoretical analyses of the various tasks are more notable for their similarities than for their differences.

Our discussion in this section is focused on empirical methods that have been used to infer the nature of procedural knowledge. First, we discuss inferences based on patterns of errors that occur in elementary arithmetic. Then we discuss inferences from latency data.

II.C.1. Diagnosis of Cognitive Procedures from Patterns of Errors. Brown and Burton (1980) analyzed children's knowledge for subtraction problems with multidigit numbers. Their data were obtained in an arithmetic achievement test taken by 1325 school children. Ordinarily, performance on tests is used to assign a simple score for each student, allowing judgments of which students have learned a satisfactory amount. Brown and Burton's analysis shows that test data are potentially much richer, and can be used to make stronger inferences about the nature of children's knowledge.

The more powerful theoretical use of test data depends on two things. First, performance on the test is not characterized simply by the number of problems correct, but by the specific answers given to all the problems, with particular attention to the incorrect answers.*3 Second, the analysis of each student's test performance consists of a model of a procedure for solving the problems.

Table 8 here

An example of an individual's performance is in Table 8. Table 8 contains six errors (the fourth problem in the second row, and all the problems in the third row), not a very good score. However, all but one of the errors were apparently caused by a single flaw in the student's procedure. When the student had to borrow and encountered a zero, the student replaced the zero by a nine, but did not go further and decrement other digits in the top number.

Brown and Burton developed a general model of subtraction for which various flawed versions can be represented as variants. The desired outcome is that the performance of each individual child, such as that shown in Table 8, should correspond as closely as possible to one of the variants of the general model.

The general model has the form of a procedural network, the formalism developed by Sacerdoti (1977) and used by Greeno et al. (1979) to explain constructions and set in geometry problem solving. The main features of a procedural network are that units of knowledge correspond to actions at differing levels of generality, and each action unit includes information about conditions for performing the action and the action's consequences.

Figure 8 here

Figure 3 shows the action components in Brown and Burton's procedural network for subtraction. The diagram shows component procedures and their subprocedures, but does not show any of the control information that is also required. For example, the diagram includes a procedure Subtract-Column, and three subprocedures, Borrow-Needed, Do-Borrow, and Complete-Column, which can be called from Subtract-Column. Control knowledge involving these subprocedures includes the information that Borrow-Needed is a test that determines whether it is necessary to borrow before finding the difference in the column, and the outcome of that test determines whether Do-Borrow will be called.

Brown and Burton formulated models of faulty performance by varying components of the procedural network for correct subtraction. For example,

*3. The idea of using patterns of errors to infer underlying psychological processes is not new, either in the psychological or the educational literature. Earlier psychological models were simpler, and the inferencers about processes were correspondingly less powerful; an example is Polson, Restle and Polson's (1965) use of errors to identify a stage of learning in which similar stimuli have not yet been discriminated. In the educational literature more complex psychological distinctions have been made, for example by Brownell (1941), but in that work, analyses of underlying psychological processes was informal, consisting of verbal descriptions of procedures hypothesized to produce observed error patterns, and as Brown and Burton documented, verbal descriptions of procedures turn out to be ambiguous in important ways.

Table 3

One Student's Performance on Subtraction Problems

(from Brown & Burton, 1973)

$\begin{array}{r} 8 \\ 3 \\ \hline 5 \end{array}$	$\begin{array}{r} 99 \\ 79 \\ \hline 20 \end{array}$	$\begin{array}{r} 353 \\ 342 \\ \hline 11 \end{array}$	$\begin{array}{r} 033 \\ 221 \\ \hline 412 \end{array}$	$\begin{array}{r} 81 \\ 17 \\ \hline 04 \end{array}$
$\begin{array}{r} 4769 \\ 0 \\ \hline 4769 \end{array}$	$\begin{array}{r} 257 \\ 151 \\ \hline 96 \end{array}$	$\begin{array}{r} 6523 \\ 1280 \\ \hline 5243 \end{array}$	$\begin{array}{r} 103 \\ 04 \\ \hline 139 \end{array}$	$\begin{array}{r} 7315 \\ 0530 \\ \hline 779 \end{array}$
$\begin{array}{r} 1039 \\ 44 \\ \hline 1995 \end{array}$	$\begin{array}{r} 705 \\ 9 \\ \hline 76 \end{array}$	$\begin{array}{r} 10038 \\ 4319 \\ \hline 15719 \end{array}$	$\begin{array}{r} 10060 \\ 98 \\ \hline 10962 \end{array}$	$\begin{array}{r} 7001 \\ 94 \\ \hline 7007 \end{array}$

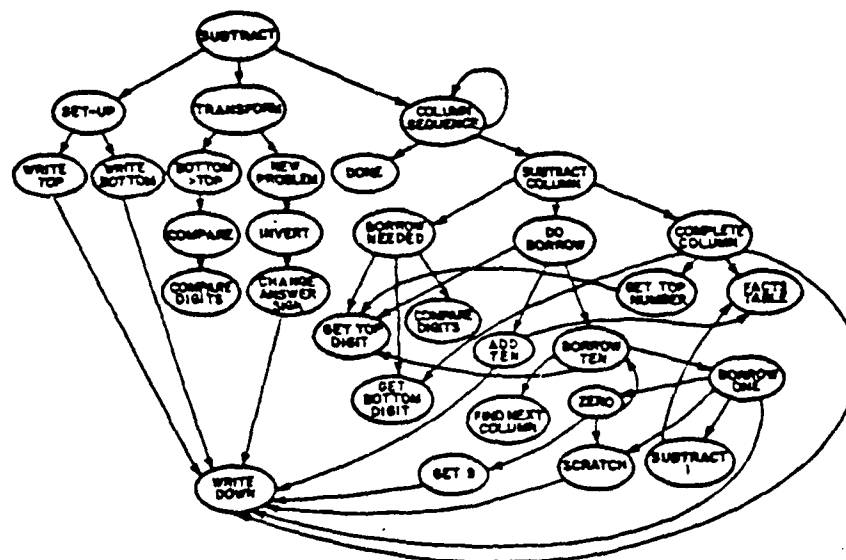


Figure 8. A procedural network for subtraction (from Brown & Burton, 1978).

the flaw of borrowing from zero is modeled by removing some of the control processing from the procedure Borrow-Ten in the Do-Borrow subprocedure. The change involves removing the decision to Find-Next-Column if a zero is found, resulting in a procedure that just changes zero to nine and adds ten to the original column.

The family of models that Brown and Burton arrived at included 60 procedural flaws of the kind described above. These 60 "bugs" provide explanations for many of the patterns of performance found in the test data, and more students' performance is explained if combinations of elementary bugs are included in the analysis. About 40% of the students' error patterns were explained reasonably well by single bugs or combinations of two elementary bugs. In examining additional sets of data, more elementary bugs have been identified (115 elementary bugs are now in the data base), and adequate explanations are typically provided for about 40% of students who make errors (VanLehn, 1982).

An alternative analysis of subtraction errors was provided by Young and O'Shea (1981), who developed a relatively simple production system that simulates correct subtraction performance, and by deleting individual productions, simulates faulty performance. Young and O'Shea's analysis provides explanations for about the same proportion of students as Brown and Burton's. On the other hand, it provides explanations for only a small proportion of the patterns of performance that have been observed. While many patterns occur rarely, their existence provides evidence for a relatively complex generative system.

Another significant development has been an effort by Brown and VanLehn (1980) and VanLehn (1983) to formulate a system that explains the production of "buggy" procedures. These formulations distinguish between a cognitive structure of partial knowledge of subtraction, and a "fallback" process of problem solving that is used when a situation is encountered for which the partial knowledge is not adequate. In VanLehn's (1983) version, the underlying cognitive structures, core procedures, result from a combination of partial learning and deletion of components of procedural knowledge. A core procedure might, for example, lack a component for dealing with a zero during borrowing. When such an impasse occurs it is assumed that the problem solver applies a general problem-solving method to be able to continue. Methods assumed to be available include skipping an operation, applying the operation to a different problem element, and using an alternative operation that is applicable in a similar problem situation. One form of evidence that supports the theory comes from data obtained by giving students repeated tests. In a substantial number of cases, students perform differently in two tests separated by two or three days, but the performance can be explained by assuming a single core procedure for which different problem-solving methods have been used.

Van Lehn (1983) conducted theoretical investigations in which a small set of problem-solving methods is combined with a plausible set of core procedures to generate buggy subtraction procedures. The generative system that has been developed can account for about one-half of the buggy procedures that have been observed; amendments that would increase the theory's empirical adequacy could be devised easily, but would not have strong theoretical motivation. Part of the progress that has been made involves identifying some general features of the system. It can be

argued, based on general properties of bugs, that the system has a push-down memory for recalling past goals, that goals are organized hierarchically, and that the representation of a goal includes the problem components to which the goal applies.

Another line of analysis that has developed from the study of subtraction bugs involves analysis of cognitive structures for understanding general arithmetic principles that underlie correct subtraction procedures. We will discuss this theoretical development in relation to the topics of representation and understanding, in Section II.D.2.

II.C.2. Inferences Based on Latencies. We now discuss an arithmetic task that is even simpler than multidigit calculation: answering basic addition problems such as $3 + 5$. The main data used in the analyses are latencies. Patterns of latencies of individual subjects are used to diagnose their solution processes.

We focus on an empirical study by Groen and Resnick (1977). Subjects in the experiment were five preschool children who knew how to count and could recognize the numerals 1 - 9, but who did not know about addition. These children were taught a method for addition using blocks. The procedure was to count out two piles, each having one of the numbers in it, and then count how many were in the two piles together. For example, for $3 + 5$, the child could count out a pile of three, then a pile of five, and then count the complete set to find eight as the answer. In showing the child the method, the experimenter sometimes started with the number on the left of the problem, and sometimes with the number on the right.

The problems used were basic addition facts involving the digits 1 - 5, omitting $5 + 5$. After a child could solve all 24 of these problems correctly using blocks, a new apparatus was introduced. The blocks were no longer provided, and the child answered problems by pressing buttons labeled 1 - 9. Children were shown how to count out answers on their fingers if this was necessary. Children received from four to seven blocks of problems with this apparatus, with about 25 problems per block.

The latency data were analyzed using regression techniques; models of cognitive processes were employed to determine the values of independent variables. Two models were used.

According to one model, the process of finding the answer to each problem was much like the procedure that the children were taught. In that procedure, a number of sets must be counted; in fact, the total number of counts equals twice the number that is the answer. If we assume that a fairly uniform amount of time is used, each time something is counted, then the total amount of time needed is

$$T = A + B(2S),$$

where S is the sum of the two numbers (i.e., the answer), and A and B are constants.

According to a second model, the process is considerably simpler. The sum can also be found by starting with the larger of the two addends and counting up the number of the smaller addend. According to this model, the time it takes to find the answer is

$$T = A + B(M),$$

where M is the minimum addend, and again A and B are constants. These two models are called the Sum Model and the Min Model, respectively.

Comparison of these two models with the data of children's performance is interesting primarily for the possibility that children spontaneously change their procedure for solving addition problems. If they use the procedure they were taught, their performance should agree with the Sum Model. However, performance consistent with the Min Model would reflect a more efficient procedure, and would indicate that children had spontaneously modified their problem-solving procedures. It would thus indicate a significant capability for discovery or invention.

To apply either the Sum or the Min Model to the data, problems are grouped according to the number of counting operations they require. Because the models specify different counting operations, they imply different groupings of items. For example, according to the Sum Model, the problems $6 + 1$, $5 + 2$, and $4 + 3$ all require the same number of operations, but these problems require different numbers of counts according to the Min Model. On the other hand, the problems $4 + 3$ and $3 + 5$ require the same number of counts by the Min Model, but are different according to the Sum Model.

If a model is approximately correct, the regression based on that model should give accurate predictions of problem latency. The criterion of fit used by Groen and Resnick was the proportion of variance, R^2 , accounted for by the regression. Higher values of R^2 indicate better agreement between the latency data and the theoretical function.

Table 9 here

Table 9 shows that about one-half of the subjects were fit better by the Min Model than by the Sum Model. Values of R^2 are shown for latency data from each block of problems except the first, in which the children were getting used to the new apparatus. Subjects 2 and 4 were fit better by the Min Model, Subject 5 was fit better by the Sum Model, and Subject 1 underwent a transition, being fit better by the Sum Model in Blocks 2 - 5, but by the Min Model in Blocks 6 and 7. Another experiment, in which practice problems were presented in a systematic order, had similar results.

The important conclusion taken from these data is that children must have discovered the procedure represented by the Min Model, since they were not taught how to add in that way. Neches (1981) has developed an analysis of learning mechanisms that can produce modified procedures, and used his system to simulate changes in counting procedures for addition problems. The main ideas in Neches' model are that redundant components of the procedure can be removed, and when there are alternative ways of reaching the same result, the easier method can be chosen. For example, in the Sum procedure, the first addend is counted, then later the process of counting the combined set includes counting the first addend as a part. Noticing this redundancy leads to removal of the initial count of the first addend from the procedure. Choice of the larger addend to initialize the procedure can be made if it is noticed that the same result is obtained

Table 9. Results of Applying Regression Models to Latency Data
(Groen & Resnick, 1977).

Subject	Block	Proportion errors	Proportion covert	R^2		Slope of best fitting line (seconds)
				Sum	Min	
1	2	.15	.02	.78*	.65*	.92
	3	.09	.00	.45*	.16	.60
	4	.03	.04	.79*	.38	.67
	5	.03	.08	.69*	.57*	.91
	6	.06	.33	.50*	.59*	1.66
	7	.05	.34	.40*	.63*	1.90
2	2	.18	.40	.44*	.65*	2.82
	3	.14	.57	.51*	.88*	2.30
	4	.11	.57	.51*	.69*	2.06
	5	.06	.76	.22	.38*	1.40
	6	.06	.99	.23	.54*	.40
	7	.11	1.00	.17	.43*	.26
3	2	.04	.00	.14	.00	—
	3	.03	.09	.71*	.57*	.99
	4	.05	.05	.50*	.27	.73
	5	.11	.30	.06	.13	—
	6	.12	.92	.05	.30	—
	7	.07	.83	.03	.10	—
4	2	.25	.73	.23	.54*	1.77
	3	.12	.61	.38*	.41*	1.60
	4	.06	.94	.32	.65*	1.38
5	2	.04	.94	.47*	.43*	1.30
	3	.09	1.00	.55*	.49*	1.66
	4	.02	1.00	.25	.12	—
	5	.01	.99	.21	.17	—
	6	.06	1.00	.52*	.20	.64

Note. Asterisks denote slope significantly different from zero at .01 level. Italics denote maximum R^2 .

regardless of which addend is used, and less effort is required if the larger addend is chosen. To produce modifications in its procedures, Neches's system requires a trace of its activity, including the goals that are active during the various stages of its performance.

The regression method also has been used in analyzing performance of adults in simple arithmetic tasks. Groen and Parkman (1972) found that college students' performance is quite consistent with the Min Model. The slope of the best fitting regression equation is far too small to correspond to verbal counting, but an analogue of a counting procedure might be postulated to account for the result.

More recent studies of performance in mental arithmetic have been conducted by Ashcraft and his associates. Using a task in which subjects are shown a problem with a possible answer and are asked whether it is correct, Ashcraft and Battaglia (1978) found an effect of problem size (i.e., longer latency when problems involve larger numbers), but this effect was not linear in the smaller addend, as required by the Min Model. A better predictor of latency was the square of the problem sum, an effect that seems inconsistent with a simple process of counting. Ashcraft and Battaglia also found that latencies were faster for rejecting wrong answers that are very different from the correct answer, compared to wrong answers that are near the correct answer. Another relevant finding by Winkelman and Schmidt (1974) is that latency was slowed by a false answer that would be correct for a different operation; for example, $3 \times 4 = 7$. As Ashcraft and Stazyk (1981) have argued, these findings can be explained most easily by assuming a process of retrieval from memory, rather than a counting procedure, with effects on latency that result from the way in which information is stored and from processes of activation and search.

II.D. Problem Understanding; Representation

Before a problem can be solved, it must be understood. Many problems used in education are presented as natural-language texts that describe situations and ask questions, usually the values of some quantities. In laboratory studies, problems often are presented in the form of instructions that specify the goals and problem-solving operators that can be used in working on the problems. These texts or instructions must be interpreted, and some kind of representation of the problem must be generated before problem-solving processes can be put to work in seeking a solution.

The same problem may be represented in radically different ways. This is illustrated dramatically by the "mutilated checkerboard" problem. We are given an ordinary 8×8 checkerboard, with alternating black and red squares, and a set of dominoes, each of which is exactly the right size to cover two squares. The entire board can be covered neatly by 32 dominoes, with no square uncovered, and no domino hanging over the edge of the board. Suppose now that the north-east square and the south-west square of the checkerboard are cut off, leaving 62 squares. Can the mutilated board now be covered neatly by 31 dominoes?

It is impossible for a human being or a computer to answer this question by exhaustive search in the obvious but enormous problem space in which the squares and dominoes are represented directly. Consider, however, an abstract problem space in which we represent only the number of dominoes that have been laid down, and the numbers of black squares and of red squares that remain uncovered. At the outset, because of the mutilation, there are 32 red squares, but only 30 black squares (or vice versa). Each domino covers exactly one red and one black square. Hence, no matter how the dominoes are placed on the board, after 30 have been placed, if that is possible, two red squares and no black squares will remain uncovered. But the final domino cannot cover two red squares, hence there is no way to complete the covering. Here, a change in problem representation changes the problem from one that is practically unsolvable to one that is solvable relatively easily.

Figure 9 here

Another famous example of problem understanding, discussed by Wertheimer (1945/1959), arises in finding the area of a parallelogram. Students are taught that the area of a parallelogram can be calculated with a formula $A = b \times h$, where b and h are the base and height, respectively. Wertheimer described two ways in which the formula may be understood. In one representation, b is the length of a horizontal side of the parallelogram, and h is the length of a vertical line drawn from a corner at the top of the figure to its base, as shown on the left of Figure 9. Many students, apparently using that representation, become confused if they are then asked to find the area of a parallelogram oriented differently, as in the right side of Figure 9. Another way to understand the formula includes a relationship between parallelograms and rectangles. A parallelogram can be transformed into a rectangle by removing a triangular piece from one end and attaching it to the other end. Then b and h are equal to the length and width, respectively, of the rectangle



Figure 9. Parallelograms in two orientations. Some students who learn the formula $A = b \times h$ have difficulty applying it to a figure like the one on the right. (Wertheimer, 1945/1951).

that the parallelogram can be transformed into. Children who understand the parallelogram problem in this way have no difficulty in solving problems where the figure is oriented differently, and frequently can transfer their knowledge to solve more complex problems, such as finding the area of a trapezoid. The two representations involve different features of specific problems, one with b and h identified with specific locations in the figure, and the other with b and h defined in more general terms.

In Section II.D.1, we discuss studies of problem understanding, involving instructions for novel problems and text problems in domains where the problem solver already has learned the problem-solving operators. In Section II.D.2, we discuss understanding of the structures of problems and problem solutions, in contrast to mere rote or mechanical knowledge of problem-solving procedures, the issue emphasized by Wertheimer and other Gestalt psychologists.

II.D.1. Understanding Problem Instructions. In most studies, consideration of subjects' behaviors in problem-solving tasks is begun after the subjects have received the problem instructions, including the definition of the problem, and have been tested by the experimenter for their understanding of the problem. In a few cases that we discuss here, the processes studied are those required for assimilating the problem prior to making attempts to solve it.

In the situations that have been studied, solution of the problem is likely to proceed by a form of means-ends analysis. Therefore, the information that subjects extract from instructions is probably similar to the information needed by the General Problem Solver. When GPS is given a problem, it is provided with a list of the objects with which the problem is concerned, the relevant properties of these objects, operators for legal moves, a description of the starting situation, and a set of tests to determine when the final goal has been reached. It must either be provided with, or acquire by learning, a set of tests for differences between situations, and a set of productions that evoke, when particular differences are present, operators that are relevant for reducing these differences.

For example, in the Tower of Hanoi problem, the objects are disks (N in number) and pegs (3). A legal move consists in transferring the smallest disk on some peg to another peg that holds no smaller disk. Hence, the size of a disk is its relevant property. Situations differ with respect to which disks are on a particular peg, or with respect to the peg on which a particular disk is located. In the starting situation, all the disks are held, say, on a single peg; the goal is to move the entire set of disks to some particular other peg. The problem description must provide all of this information, in English, and the subject (or computer program) must convert this English prose into an internal representation that permits situations and moves and their consequences to be modeled. A disk, for instance, may be represented as a schema, one of whose attributes is its size; a peg by a schema, one of whose attributes is the list of disks currently on that peg. A move operator is a process that changes a pair of the latter lists, by moving the name of a particular disk from the one list to the other.

Two central problems for psychological research on the understanding of problem instructions are: (1) how the verbal instructions are converted to an internal representation; and (2) what characteristics of the instructions cause the problem to be represented in one way, rather than other possible ways. The second question is especially important when alternative representations differ in the difficulty of solving the problem, as with the mutilated checkerboard example, or provide differing degrees of generality, as with the parallelogram problem. The questions have been addressed by Hayes and Simon (e.g., 1974), who obtained information about internal representations by collecting extensive verbal protocols of problem-understanding processes. By using problems where alternative representations were available, Hayes and Simon have also cast light on the question of which representation will be formed.

The Understand program (Hayes & Simon, 1974) is a computer simulation of the problem understanding process for puzzle-like problems like the Tower of Hanoi or the Missionaries and Cannibals problem -- that is, for problems that do not assume the subject has any prior knowledge of the problem domain. The program matches human thinking-aloud protocols sufficiently well to lay claims of being a good first-approximation model of the process.

Understand operates in two principal phases. In the first phase, a language-parsing program extracts the deep structure from the language of the instructions. In the second phase, another set of processes constructs from this information a problem representation that is suitable as input to a GPS-like problem-solving program. This is accomplished by (a) identifying the objects and sets of objects that are mentioned in the parsed text, (b) identifying the descriptors of those objects and the relations among them, (c) identifying the descriptions of legal moves and constructing move operators that fit those descriptions, (d) identifying the description of the solution and constructing a test for attainment of the solution, and (e) constructing an organization of schemata that describes the initial problem situation.

For example, after parsing the written description of the Tower of Hanoi problem, Understand would identify pegs and disks as the relevant sets of objects, and would notice that disks are on pegs and that disks move from one peg to another. It would extract the information that only the smallest disk on a peg may be moved, and only to a peg where there is no smaller disk, and would construct a test process for checking these conditions. It would determine that the problem is solved when all the disks are on (say) the third peg, and would construct a test to determine when that condition is satisfied. Finally, it would generate a list structure showing all the disks initially as being on the first peg. From the evidence of protocols, and of subjects' subsequent problem-solving behavior, this is what human solvers do also.

Problem Isomorphs. A powerful experimental manipulation for studying problem understanding is to use variant problem instructions all of which describe isomorphs of a single problem. Two problems are isomorphs if the legal problem situations and legal moves of the one can be mapped in one-to-one fashion on the situations and moves of the other. Then, if situation S' is the isomorph of S and moves A' , B' , etc., are the isomorphs of A , B , etc., and if the succession of moves A, B, \dots takes the one

system from S to T, then the succession of moves A', B',... will take the other system from S' to T', where T' is the isomorph of T.

Using a number of isomorphs of the Tower of Hanoi problem, Hayes and Simon (1977) demonstrated that problem difficulty varied by a factor of two to one from one class of problem descriptions (transfer problems) to another class (change problems). Moreover, protocols and diagrams produced by subjects showed that they were consistently using different representations for the different classes of isomorphic problems. The Understand program behaved in the same way, constructing different representations for the transfer and change problems, respectively. In only one case out of the nearly 100 that have been examined did a subject shift from the more difficult "change" representation to the easier "transfer" representation.

The reasons why the change problems take twice as long to solve as the isomorphic transfer problems have not been fully elucidated. It can be shown, however, that the tests for move legality are a little more complex for the former than for the latter, and this additional complexity may increase the short-term memory load on the subject who is seeking to understand the problem instructions.

Problem isomorphs can be used to study transfer of training, and such a study was conducted by Reed, Ernst, and Banerji (1974). They devised a variant of the missionaries and cannibals problem, called the jealous husbands problem. The latter differs from the former in that specific husbands are paired with specific wives, and no woman may be left in the company of men unless her husband is present. Experimental results showed that subjects were not better at solving one of these problems if they had previously solved the other. We must conclude that, while subjects may use analogies to help solve problems, there is nothing automatic about the availability of an analogy, and subjects may fail to take advantage of analogies unless their attention is drawn to them or they are made salient in some other way. (Experimental results showing positive transfer between problem isomorphs for a somewhat different type of problem are discussed in Section III.A.3.)

II.D.2. Problem Representation in Mathematics and Physics.
Typically, a problem given in a mathematics or physics text describes a situation, including quantitative values of some variables, and asks for the value of another variable. The given quantities correspond to the initial state of a problem and the unknown quantity provides the goal. The problem is presented in a natural-language text as are the instructions for novel problems that we discussed in the previous section. The situation with a physics or mathematics problem differs from a puzzle in that the instructions for the former do not provide a description of the problem-solving operators that can be used. The student is assumed to already know the operators, based on class instruction or reading the text. The interpretation of puzzle instructions is a representation that can be used by a general problem-solving system such as GPS, while the interpretation of a text problem in mathematics or physics is a representation that can be used by domain-specific problem-solving procedures.

Algebra Word Problems. Word problems in algebra describe situations that can be translated into equations, which are then solved to find the values of unknown variables. An early model of solution of word problems called Student (Bobrow, 1968) showed that the translation can be accomplished mainly using the forms of sentences in the problem text, and of course the numerical quantities, with very little knowledge about the objects that are described. For example, in the sentence "The number of customers Tom gets is twice the square of the number of advertisements he runs," Student does not need to know anything about what customers or advertisements are, but can form the equation $X = 2Y^2$ using the function words "is" and "of" in critical ways.

In an empirical study of the solving of algebra word problems, Paige and Simon (1966) found a good deal of similarity between human solutions and those given by Bobrow's Student program. However, they found that their more skillful subjects used an intermediate semantic representation in the translation of the English-language problem statements into algebraic equations. Some problems presented descriptions of situations that were contradicted implicitly by real-world knowledge (boards of negative length, nickels worth more than quarters, and so on). The weaker subjects often made accurate syntactic translations of English into equations, as Student does, even though the equations represented nonsense situations. The abler subjects either noticed the contradictions between the statements and their knowledge, or translated the statements (carelessly) into equations that were not quite equivalent syntactically, but which represented physically realizable situations.

Another difference between the abler and weaker subjects was that the former, but not the latter, generally drew diagrams of the problem situation that contained all the essential relations from which the equations could be derived.

Both kinds of evidence -- the response to "impossible" situations, and the nature of the problem diagrams produced -- point strongly to the employment by the more competent subjects of an intermediate semantic representation of the problem situations, rather than a direct translation from English to algebra.

Arithmetic Word Problems. Detailed analyses of intermediate representations have been worked out for a class of word problems in elementary arithmetic. Riley, Greeno, and Heller (1983) and Briars and Larkin (in press) have developed models of representation and solution of word problems that are solved by a single operation of addition or subtraction. Examples of the problems studied are "Jay had eight books; he lost five of them; how many books does Jay have now?" or "Jay has some books; Kay has seven more books than Jay; Kay has eleven books; how many books does Jay have?"

In Riley et al.'s (1983) model, problems are represented by three schemata that provide knowledge of basic quantitative relationships. One schema represents problems involving events that change the value of a quantity, either by increasing it or decreasing it, as with losing five books. A second schema represents problems in which two separate quantities are considered in combination. A third schema represents problems involving comparison between two separate quantities. (This

classification of problems is not unique; similar but distinct characterizations have been worked out by Carpenter and Moser, 1982; by Nesher, 1982; and by Vergnaud, 1982.)

Arithmetic word problems are usually classified according to the operations used in their solution, and children are often taught to look for certain key words to decide how to solve the problems. This is inadequate, because choice of the correct operation depends on understanding the structure of quantities in the problem, rather than on a single feature corresponding to a key word. For example, "altogether" is sometimes suggested as a key word for addition, but this is not a reliable cue, as in the problem "Jay and Kay have nine books altogether; Jay has seven books; how many books does Kay have?"

Riley et al.'s model simulates children's solutions of word problems when small blocks are available for the children to use in solving the problems. The model forms representations of problem texts using the schemata of change, combination, and comparison. Based on the representation that is formed for a problem, the model performs quantitative actions, such as joining two sets of objects together or taking a specified number of objects away from a set and counting how many remain. Different versions of the model were formed to correspond to different levels of skill that were observed in a study of children from kindergarten through the third grade. The versions differ in the detail with which internal representations are formed (affecting ability to retrieve information from earlier steps), and in their ability to perform transformations that provide information in a form needed to make inferences. The patterns of correct responses and errors observed in the performance of most of the children were consistent with patterns that were obtained in the simulation models.

Briars and Larkin's (in press) model constructs less elaborate intermediate representations of problems, and thus relies somewhat more strongly on procedures for inference. Briars and Larkin's model does use a schema for representing part-whole relations among sets for some relatively difficult problems.

Physics Problems. The knowledge structures used in simulating solutions of arithmetic word problems are quite general, involving relationships between quantities that children probably learn about in their ordinary experience. In technical domains such as physics, specific instruction is given to teach students the nature of theoretical quantities and the ways that they combine.

Novak (1977) constructed a program called Isaac that builds problem representations from English problem descriptions in a domain of physics, simple statics problems. Isaac uses schemata of physical subsystems (levers, masses, etc.), assumed already understood by the solver, to build up a compound schema to fit the problem at hand. Thus, it may assemble a wall schema (surface), a floor schema (surface), a ladder schema (lever), and a man schema (mass) to represent a situation of a man standing on a ladder that is leaning against a wall, assigning to each component appropriate numerical quantities and appropriate connections with the others.

Models such as Riley's for arithmetic word problems and Novak's for physics problems are based on the idea that understanding a problem requires schematic knowledge regarding the quantities in problem situations. The schemata provide knowledge of ways in which quantities are related to one another. These quantitative relations are not expressed adequately in algebraic formulas that are taught in physics and other quantitative sciences, although of course the formulas are based on the quantitative relations and students must be able to choose formulas and assign values of their variables correctly on the basis of the problem representations that they construct.

The distinction between knowledge of a formula and knowledge of quantities and their relations is illustrated in experiments conducted by Mayer (1974). The experiments were instructional studies, concerned with different methods of teaching the formula for binomial probability. One group of subjects received instruction that emphasized calculation, presenting components of the formula with explanations of the calculation steps, practice exercises, and relatively brief explanations of the referents of terms in the formula. Another condition emphasized the information needed for students to acquire schematic knowledge, presenting definitions of terms and explanations of relevant concepts such as the number of combinations and the probability of a single sequence of outcomes before calculation exercises were presented. Tests that were given following instruction included a variety of problems, including some that involved direct application of the formula, and others that required more interpretation. The latter group included word problems, problems that could not be solved because of inconsistent or insufficient information, and problems requiring use of a component of the formula rather than the complete formula. The subjects whose instruction emphasized the formula excelled on the problems involving direct use of the formula, but the subjects given more conceptual instruction were more successful on the problems requiring more interpretation.

Several studies have compared performance of physics students with performance of expert physicists to identify some of the components of knowledge that characterize more advanced problem solvers. Three of the differentiating characteristics that have been identified are (1) use by experts of abstract physics principles in representing problems as well as in providing methods of solution; (2) strong organization of physics knowledge including knowledge of relationships among principles and recognition of complex patterns of problem features; and (3) integration of physics knowledge with general concepts and reasoning processes.

Experts' use of abstract physics concepts was shown in experiments by Chi, Feltovich and Glaser (1981), who gave subjects a set of 24 physics textbook problems and asked the subjects to sort the problems into groups. Groupings formed by advanced graduate students were based primarily on abstract principles, such as conservation of energy, while groupings formed by subjects who had completed a single course in mechanics were much more likely to be based on superficial features, such as the kinds of objects (pulleys, levers, etc) that were mentioned in the problems. Chi et al. also found that abstract physics principles were used by experts when they gave protocols reporting their thoughts and hunches while deciding on a "basic approach" to solving the problem. Use of abstract principles was included in a computational model developed by McDermott and Larkin (1978)

which simulates representation of textbook problems by an expert. The representation of a problem includes a diagram that represents major components and relations, and then an abstract representation with theoretical entities such as forces and energies and relations among these based on general principles.

Instructional materials have been designed by Reif and Heller (1981) that provide training for beginning students in a procedure for constructing abstract representations of problems. Reif and Heller's instruction provides an explicit method for arriving at a correct problem representation like that used by experts (although the representational method is not patterned after the experts' performance, in which the process of forming the representation is rapid and apparently automatic, without easily discerned intermediate steps).

Larkin and Reif (1979) also designed instruction to strengthen students' knowledge of relations among physics principles and their ability to apply principles in solving problems. The instruction grouped principles on a chart, and suggested to students that when certain principle are applied, it is generally useful to consider the application of other related principles. Qualitative analogies were also used, such as a fluid-current analogy for electric current and a height analogy for potential. Students who received this instruction were more successful in solving test problems than other students who only received instruction in the principles, without the organization and qualitative analogies.

Individuals with expert knowledge in a domain have been shown to have superior skill in recognizing complex patterns of information in their domain of expertise. Domains in which this phenomenon has been demonstrated include chess (Chase & Simon, 1973), Go (Reitman, 1976), electronics (Egan & Schwartz, 1979), computer programming (McKeithen, Reitman, Ruiter, & Hirtle, 1981), and radiology (Lesgold, Feltovich, Glaser & Wang, 1981). (We discuss experiments on chess perception in Section III.B.2.) Highly developed skill in pattern recognition may provide an explanation for a finding that has been obtained in several studies, namely that expert problem solvers tend to work forward from the given information to the unknown, while novices tend to work backward from the unknown, searching through a series of subgoals for formulas that can provide the needed quantities (e.g., Simon & Simon, 1978). The conditions for applying formulas involve relatively complex patterns of known values of variables, which experts probably have learned to recognize directly, thus avoiding the more laborious searches that novices conduct (Larkin, 1981). A result supporting this view was obtained by Malin (1979), who found that subjects were more likely to adopt a forward-search strategy to solve problems if the formulas that they were using had an obvious organization than if the formulas did not fit together in any evident way.

A third characteristic of experts' knowledge is that their domain-specific knowledge (e.g., in physics) is integrated with powerful general concepts and procedures for making inferences. An example is in protocols obtained by Simon and Simon (1978) from a novice and an expert subject on problems from a high school physics text. One problem was: "An object dropped from a balloon descending at four meters per second lands on the ground ten seconds later. What was the altitude of the balloon at the moment the object was dropped?" The novice subject's solution had the

properties of means-ends analysis, using the formula $s = v_0t + .5at^2$. In contrast, the expert calculated a quantity that he called "the total additional velocity" by multiplying the time by the gravitational constant (i.e., $10 \times 9.8 = 98$), added that to the initial velocity to obtain the final velocity ($98 + 4 = 102$), took the average velocity $((4 + 102)/2 = 53)$, and found the distance by multiplying the average velocity by the time of ten seconds ($53 \times 10 = 530$ meters). The expert apparently had a representation of the problem in terms of physical quantities that enabled him to apply general procedures such as computing components of velocity and taking an average, whereas the novice was restricted to using the formulas that were provided. Relations between technical knowledge and general concepts have been investigated theoretically by deKleer (1975) and by Bundy (1978), who developed models of physics problem solving that combine general knowledge about the motion of objects on surfaces with knowledge of formulas in kinematics, and by Larkin (1982) who studied the use of spatial information in solution of hydrostatics problems.

Understanding of Structure and Principles. Integration of problem-solving knowledge with general conceptual structures also has been used to characterize structural understanding, like that discussed by Wertheimer (1945/1959), and understanding of general principles, including the relation of abstract properties of number (cardinality, order, one-to-one correspondence) to children's cognitive procedures for counting.

Understanding of structure has been investigated theoretically by Greeno (1983) using a problem discussed by Wertheimer (1945/1959), proof of the congruence of vertical angles. Wertheimer distinguished between a relatively mechanical process for generating the proof, involving use of algebra without cognizance of spatial relations in the problem, and a more meaningful process based on part-whole relations between pairs of angles and operations of removing a part that is included in two whole angles. Greeno's model simulates the more meaningful process by using a schema that represents part-whole relations in a general way and applying problem-solving operators that make inferences based on the part-whole structure. Data were available in the form of protocols from students working on the vertical-angles problem after they had learned to solve other problems with similar part-whole structure involving line segments. The model simulates learning in the line-segment situation. When the learned problem-solving operators are integrated with the part-whole schema, the model can apply its knowledge when it encounters the vertical-angles problem. The model thus provides an explanation for transfer that occurs between problems in different domains, with a characterization of structural understanding based on schematic representation. An account of transfer based on acquisition of a schema in a different problem domain is discussed in Section III.A.3.

A similar idea was used by Resnick, Greeno and Rowland (described by Resnick, 1983) in analyzing children's understanding of a procedure for subtraction with multidigit numbers. According to their analysis, children who understand the procedure have a representation that includes general relations, such as part-whole relations between quantities represented by individual digits and the quantities represented by whole numbers, and constraints such as requiring the quantity represented by a number to remain unchanged when borrowing is used. The analysis focused on knowledge

acquired in meaningful instruction (cf. Brownell, 1935), in which children were shown the correspondence between subtraction with numerals and an analogous subtraction procedure using blocks. Resnick et al. hypothesized that the understanding was achieved through acquisition of a schema, involving general part-whole relations, that was general enough to apply to both of the domains: the numerals and the blocks.

Efforts also have begun to develop rigorous and explicit characterizations of knowledge that includes implicit understanding of general principles (cf. Judd, 1908; Piaget, 1941/1952). A representation of preschool children's understanding of principles of counting has been formulated by Greeno, Riley, and Gelman (in press). Greeno et al.'s analysis was based on evidence presented by Gelman and Gallistel (1978) that young children have significant understanding of principles such as cardinality, order, and one-to-one correspondence, rather than simple "mechanical" knowledge of counting procedures. The evidence includes performance in novel situations, such as being asked to evaluate counting performance by a puppet that sometimes makes errors, and counting with a novel constraint of associating a specified numeral with a specified object. Greeno et al. proposed an analysis of conceptual competence to represent children's implicit understanding of principles. Conceptual principles are represented as schemata that incorporate constraints on correct counting and express general properties such as the part-whole relation between the counted objects and the whole set. The conceptual principles are related to procedures of counting by a set of planning rules, which permit derivation of procedures from the schematic representations of the principles.

III. Problems of Design and Arrangement

Problems that we discuss in this section require finding an arrangement of some objects that satisfies a problem criterion. Simple examples include puzzles in which the objects are given in the problem situation. For example, an anagram presents some letters, and the task is to find a sequence of those letters that forms a word. In more complex cases, the problem solver provides the materials based on his or her knowledge. Examples are composing an essay and writing a computer program.

The problem space in a problem of design includes the objects that are given or are in the problem solver's knowledge. The space of possible solutions is the set of arrangements that can be formed with the available objects. The problem goal is to construct an arrangement that meets a criterion, which may be either specific or nonspecific. An anagram problem has a specific criterion: the sequence of letters should form a word. A written composition has several criteria that are less specific, such as clear exposition, persuasive argument, and an entertaining style. Many problems of design have a mixture of specific and nonspecific criteria. For example, a problem in computer programming can have a criterion of a specific function to be computed, and less specific criteria such as efficient computation and clarity of structure.

An important factor in solving problems of design is the satisfaction of constraints. The metaphor that best characterizes typical solution processes is "narrowing down the set of possibilities" rather than "searching through the set of possibilities." Although it is entirely possible, as we shall see, to describe the solution process as a search, the main steps in this search lead to the acquisition of new knowledge that rules out a whole set of problem states as potential solutions -- a wholesale approach to the reduction of uncertainty. Use of constraints is important because the set of possible arrangements is usually very large, compared to the set of arrangements that satisfy the problem criterion.

Problems of design are differentiated from transformation problems, discussed in Section II, in the nature of the goal and the set of alternatives that are considered. In a transformation problem such as the Tower of Hanoi or finding a proof for a theorem, the goal is a specific arrangement of the problem objects, such as a specific location of all the disks in the Tower of Hanoi or a specific expression to be proved in logic. Thus, the question is not what to construct, as it is in a problem of design, but how the goal can be constructed with the limited set of operators that are available. The search for the solution of a transformation problems often examines one problem situation after another, gaining knowledge that helps point the direction of the search toward the goal situation.

Viewed in another way, however, transformation problems and problems of design are very similar in structure. The solution of a transformation problem is a sequence of actions that changes the initial problem situation into the goal. The solution process can be considered as construction of an appropriate sequence of actions, involving search in the very large space of possible sequences. This view emphasizes similarities between problems of transformation and problems of design, which are especially apparent when solution of transformation problems includes planning.

This section discusses design problem solving in four parts. In Section III.A we discuss two simple problems of forming arrangements: cryptarithmic and anagrams. These provide paradigmatic cases for the analysis of problems of search among sets of possible arrangements. Section III.B discusses problems in which an arrangement of objects is already presented, and the task is to modify the arrangement according to some criterion. These problems include classical studies of matchstick puzzles by Katona (1940), and selection of moves in board games where the goal of a player is to strengthen his or her position. In Section III.C we discuss so-called "insight" problems, which depend on finding a successful formulation or representation of the problem. Finally, Section III.D discusses several more complex problems of composition and design, including composition of essays and musical pieces, design of procedures, and formation of administrative policies.

III.A. Simple Problems of Forming Arrangements

First, we discuss cryptarithmic problems, analyzed by Newell and Simon (1972), in which digits are arranged to form a correct addition problem, constrained by a set of letters for which the digits are to be substituted. Then we discuss anagram problems, where letters are to be arranged to spell a word.

III.A.1. Cryptarithmic Problems. Cryptarithmic problems are best explained by exhibiting one of the best known examples:

$$\begin{array}{r} \text{D O N A L D} \\ + \text{G E R A L D} \\ \hline = \text{R O B E R T} \end{array}$$

The task is to replace each distinct letter in the array with a distinct digit, from 0 to 9, the same digit replacing a given letter in all its occurrences, and no digit being used for more than one letter. To make the problem easier, the solver is usually told that $D = 5$.

The cryptarithmic task was apparently first studied by Bartlett (1958), who reported some retrospective protocols of subjects in his book on thinking. Subsequently, Newell and Simon (1972) carried out extensive analyses of thinking-aloud protocols for cryptarithmic problems. From this work, we now have a rather clear picture of how human subjects approach such problems.

There are $10! = 3,628,800$ ways of assigning ten digits to ten letters. Most subjects, without calculating this number, realize that it is very large, and do not even attempt to solve the problem by making random assignments and testing them. Instead, they look for information in the form of constraints that permit values to be assigned to particular letters at once. If that can be done, the number of possibilities declines rapidly. Simply giving the information that $D = 5$ already reduces the possible solutions by a factor of ten, that is, to 362,880 -- still a large number!

The constraints in cryptarithmic problems that sometimes make systematic elimination possible derive from the fact that each column of the literal array must be translated into a correct example of addition (subject to carries into and out of the column). Thus, as soon as it is

known that $D = 5$, the sixth column can be processed to infer that T necessarily equals 0, and that there is a carry of 1 into the fifth column. This single inference reduces the remaining set of possible assignments to 40,320, a further reduction by a factor of nine.

Similarly, consideration now of the second column allows the subject to infer that E is equal to 0 or 9. Since 0 has already been preempted by T , we have $E = 9$, reducing the possible assignments to 5,040. A few more steps of reasoning, based on information contained in columns one and five, allow the subject to infer that $R = 7$, reducing the possible assignments to 720. An inference on column four gives $A = 4$ (120 possibilities remain); and an inference on column five gives $L = 8$ (leaving only 24 possibilities). From column one, $G = 1$ (leaving 6 possibilities), and now the remaining digits must be assigned to N , O , and B , a task easily carried out by trial and error.

Newell and Simon (1972) obtained thinking-aloud protocols of subjects solving cryptarithmic problems. Problem behavior graphs were constructed, based on the protocols, and a detailed model of one subject's problem-solving processes was developed in the form of a production system. (A discussion of this methodology was given in Section II.A.1.) The model includes several productions that represent a problem-solving strategy. These productions set goals to examine a column or to examine occurrences of a variable, make decisions that a value can be assigned to a variable or that a candidate value should be tested, and perform other general functions. There also are a few dozen productions that represent the operation of specific processes. One, called Process Column, contains 26 productions; others are considerably simpler. The productions in this process examine the letters in a column and use any information that has been gathered about them to make further inferences. The subject's performance, recorded in a problem behavior graph, was compared in detail with the model, and approximately 80% of the protocol units were explained by processes in the model.

Protocols obtained from five subjects were consistent in their general characteristics of problem-solving processes. They also revealed significant individual differences, and these can be interpreted as differences between the problem spaces of the individual problem solvers. All subjects made use of their knowledge of arithmetic in order to make inferences, and subdivided the problems into subproblems involving the columns. There were important differences among subjects in their strategies for selecting columns to work on, and in their use of specific constraints for making inferences.

For an efficient solution of this problem, subjects must use a search heuristic of attacking the most constrained columns first, for the most information can be extracted from a column in which the assignment of one or more letters has already been made, or in which the same letter occurs twice. Some subjects used this column selection heuristic immediately; others began by attacking the columns systematically, from right to left, and only later abandoned that strategy for the more powerful one. Subjects who did not use the heuristic usually failed to solve the problem.

Another factor that influenced success was use of specific constraints. The problem spaces of some subjects included rules of parity. For example, one of the inferences needed to conclude that $R = 7$ is that whatever R 's exact value, it must be an odd number. This is inferred by processing column 5, containing two L 's whose sum must be even, and the carry of 1, making the total an odd number. Subjects whose problem spaces did not include the parity constraints generally were unable to solve the problems that they worked on.

Even subjects who used the available heuristics and constraints for efficient elimination found the DONALD + GERALD problem difficult. Most of their difficulties arose from one or both of two sources. One such source is the making of conditional assignments (e.g., "suppose that L is 1"). Then, if the assignment was wrong and they arrived at a contradiction, they may have been unable to remember which prior number assignments they had inferred definitely and which they had postulated conditionally. Another source of difficulty involved mistakes made in drawing inferences, resulting in incorrect assignments. For example, some subjects drew from the fact that $R = 7$ the conclusion that $L = 3$ (with a carry from the sixth column), ignoring the possibility that L might be 8, with a carry into the fourth column. When $L = 3$ led to a contradiction, they often had great difficulty in discovering the cause.

Errors of inference are forms of the errors of syllogistic reasoning that we discuss in Section V. In the particular example just cited, subjects appeared to infer from the premise "if $L = 3$ then $R = 7$ " and the premise " $R = 7$ " the conclusion, " $L = 3$," an example of the classical fallacy of inferring the antecedent from the consequent. They did not notice that $L = 8$ also implies $R = 7$. Thus, the cryptarithmic task draws upon reasoning processes as well as upon search processes.

Nothing in the behavior of subjects solving cryptarithmic problems suggests that they decide consciously to treat it as a "constraint" problem rather than a "search" problem. In fact, their behavior can be described as a search through the space of possible assignments, and Newell and Simon's analysis used this point of view. What distinguishes it from search in many other problem spaces is that the problem is factored into ten separate but interdependent searches for the individual assignments. Success in each one of these searches constrains the problem space by reducing the number of alternative possibilities for the remaining assignments, and by providing additional information about some of the columns. Hence, it is not dissimilar from an ordinary search where each step of progress provides clear feedback of information that the right track is being followed.

III.A.2. Anagrams. Anagrams are strings of letters that can be rearranged to form words, for example, thgli \rightarrow light. The problem space of an N -letter anagram contains $N!$ possibilities, hence increases rapidly with N . The solution process can be viewed as a search through this space of permutations of the letters, but most persons presented with an anagram use various heuristics to speed up the search. One of these is to pick out initial combinations of letters that are pronounceable (e.g., ti or li in the example above), and then try to complete a word with the remaining letters. Imposing the condition of pronunciability upon solution attempts may restrict the search space considerably.

The course of the search also is much influenced by the structure of long-term memory. For example, if there are two possible solutions for an anagram, the one corresponding to the more frequent and familiar word is likely to be found by a large majority of subjects. Moreover, the solution may be primed by presenting the word to the subject, or a semantically related word, some time before the anagram task is taken up (Dominowski & Ekstrand, 1967).

Perceptual factors may affect performance on anagram tasks. Anagrams that are already words (e.g., forth → froth) or are easily pronounced (e.g., obave → above) take longer to solve than those without these properties (Beilin & Horn, 1962). This finding is consistent with Gestalt principles that "meaningful" forms resist restructuring. Gavurin (1967) found a correlation of .54 between success in solving anagrams and scores on a standard test of spatial abilities. When the subject was provided with tiles that could be rearranged physically, this correlation disappeared, indicating that the original relation had to do with perceptual abilities to operate on visual or auditory images.

It is easy to induce a problem-solving set in anagram solving by presenting subjects with anagrams that all represent the same permutation (say, 5 4 1 2 3) of the letters. If an ambiguous anagram (one admitting several solutions) is then presented, most subjects will find the solution requiring the same permutation rather than the alternative solution (Rees & Israel, 1935).

Thus, we find in subjects' behaviors on the anagram task a combination of search (generating possible solutions) and constraint satisfaction (rejecting non-pronounceable initial segments). The process of alternative generation, in turn, is strongly influenced by long-term memory organization and priming, and by subjects' skills in forming and holding in STM the permutations of the stimulus.

III.B. Problems of Modifying Arrangements

In the problems discussed in Section III.A, arrangements are formed "from scratch;" materials are provided, and the problem solver must put them together in a way that satisfies a specified criterion. Now we discuss problems in which an arrangement of objects is presented, and the task is to modify the arrangement. We will discuss a laboratory problem of this kind, the matchstick problem studied by Katona (1940), and problems of choosing moves in board games such as chess and Go. Perceptual processes play an important role in solution of these problems, which involve recognition of general features and complex patterns in the arrangements that function as cues for the selection of operations and strategic plans.

These problems combine features of the transformation problems discussed in Section II with features of problems of design. Like design problems, a goal is specified as a general criterion rather than a specific state that the problem solver tries to produce. At the same time, in these problems there are significant restrictions on the operators that can be used to change the situation. Therefore, the problems can be conceptualized either as search in a space of possible arrangements or in a space of possible sequences of moves.

Figure 10 here

III.B.1. Matchstick Problems. One of the matchstick problems used by Katona (1940) is shown in Figure 10. The 16 matches form five squares; the task is to move exactly three matches in such a way that the matches form only four squares, but all matches are used as sides of squares. Katona tested subjects under three conditions: a rote learning condition (subjects were shown a specific solution, and required to learn it), a condition in which a logical condition for the solution was taught (that in the solution, each match was a side of one and only one square), and a condition in which a heuristic for solving the problem was taught ("You need to open up the figure").

The subjects learned the solutions, then were tested on transfer tasks (different initial arrangements of the matches and different numbers of squares). Two weeks later they were invited back and tested for their memory of the solution. Differences in the ease of learning the solution were minimal, with the rote solution being learned most rapidly. With respect to transfer and retention, however, the "logical" and "heuristic" solutions far outshone the rote solution, and the heuristic solution scored slightly better than the logical. Katona concluded from this evidence that problem-solving knowledge and skills are better available for transfer and better retained when the learning is meaningful than when it is rote.

The experimental manipulations leave implicit, however, the theoretical import of the term "meaningful". Why does meaningful learning facilitate retention and transfer, and why is the "heuristic" form of the instruction somewhat superior to the "logical" form?

With respect to transfer and retention, meaningful learning involves the same issues as structural understanding, discussed at the end of Section II.D.2. Transfer is facilitated because with more meaningful instructions, subjects acquired knowledge that could be applied more generally -- in particular, to the new problems presented in the test as well as the problems used in training. It is easy to see why this would occur; the meaningful instructions can be applied to matchstick problems generally, while a specific solution sequence only applies to a single problem.

With respect to retention, it may be that meaningful forms of instruction provide more redundancy, hence more opportunity to recover from partial forgetting. The general principles of single vs. double function and of loosening or condensing the figure are constraints that can be used to limit search for information in memory, or to reconstruct solutions that are only partially remembered.

The difference between the two meaningful procedures appears to derive from the distinction between generators and tests. The instruction to "open up the figure" provides a constraint for selection of an operator -- it suggests something to do, however vague, related to a general property of the figure that can be perceived. The rule, "each match must be a side of one and only one square," is a constraint on solution arrangements. It provides a test that can be applied to an attempted solution, but does not

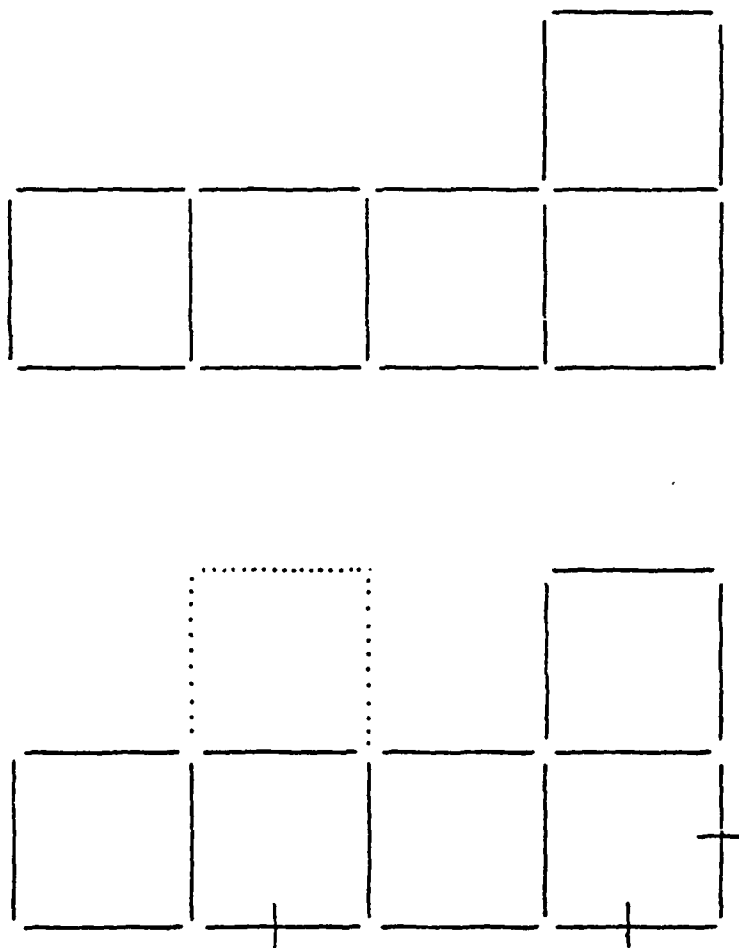


Figure 10. A matchstick problem used by Katona (1940).

suggest what attempt to make to produce the solution in the first place. In fact, the matches that are actually moved to solve the problem are not the double-function matches but matches that already lie on the side of only one square. In this situation, at least, the knowledge that facilitates solution more effectively increases the selectivity of the move generator, rather than the selectivity of candidate solution states.

Katona noted that the heuristic of opening the figure or closing gaps uses a feature that is important in the perception of form, the Gestalt principle of good continuation. Attending to that feature and considering moves to change an arrangement with respect to it constitutes a general strategy for solving matchstick problems.

III.B.2. Chess and Go. Board games present problems that have the same general form as matchstick problems. An arrangement of objects is presented -- the current situation in the game -- and a player has the task of selecting a move or a move sequence. Sometimes the criterion for a good solution is quite definite (e.g., "white to mate in four moves"); more often, it is quite general, involving a goal to achieve a stronger position. Recent experiments have compared the performance of individuals who differ in their level of skill. The results of these experiments show the importance of knowledge for recognizing large numbers of complex patterns that occur during games. The relation of this complex perceptual knowledge to strategies of play has been analyzed in recent theoretical studies.

In complex games, as in other domains in which individuals become expert, problems that would be difficult or impossible for novices are often solved "instantly" by experts -- that is, in a few seconds. For example (deGroot, 1965), when a chess grandmaster is presented with a position from an actual game with which he was previously unfamiliar, and is asked to recommend a move, he will usually be able to report a good move, often the best move, in five seconds or less. In a "blitz" game, where he is required to move within ten seconds, he will probably be unable to play grandmaster-level chess, but will achieve master level. With a high level of success, he will be able to play 50 or more opponents simultaneously, taking only a few seconds for each move. When experts are asked how they solve problems so rapidly, they may reply, "I use intuition," or "I use my judgment."

The nature of this "intuition" and "judgment" has been clarified by experiments on skill in chess by deGroot (1965) and Jongman (1968), repeated and extended by Chase and Simon (1973), and on skill in Go by Reitman (1976). A chessboard with a position from a game (containing perhaps 25 pieces) is shown to subject for five to ten seconds. The subject is then asked to reconstruct the position. Chess grandmasters and masters can perform this task with 90 per cent accuracy. Ordinary players can replace only five or six pieces correctly (20 to 25 per cent accuracy). In a second condition the task is the same, except that the pieces are now arranged on the chess board at random, rather than in a pattern that could have arisen in a game. In this condition, the performance of masters falls to the level of that of ordinary players -- both can replace, on average, only about six pieces. This second part of the experiment demonstrates that the chess masters do not have any special powers of visual imagery.

Reitman's (1976) study of skill in Go had similar results. Go is a game of territory played on a 19X19 grid. The pieces are round "stones" that are all the same except, of course, that they are different in color for the two players, black and white. An experienced subject, very strong for a non-Oriental but not as strong as a professional player, was able to reproduce 66% of the pieces of meaningful patterns, compared to 39% for a beginner who had played about 50 games. On random patterns the players replaced 30% and 25%, or an average of five to seven stones.

(This experimental procedure has now been used to study pattern-recognition abilities of experts in several other domains; see our discussions of programming in Section III.D and electronics in Section IV.D.)

The behavior of the chess and Go experts in the perception and memory task can best be explained as a function of their chess and Go experience. As a result of thousands of hours of time spent in looking at game boards, they are familiar not only with the individual pieces, but with many configurations of three, four, or more pieces that recur time and again in games. For example, a configuration known as a "fianchettoed castled Black King's position" occurs in about one out of every ten games between expert chess players. This configuration is defined by the positions of six pieces. It has been estimated (Simon & Barenfeld, 1969; Simon & Gilmarin, 1973) that a chess master has stored in long-term memory not fewer than 50,000 familiar patterns of this kind. This number is comparable to the 50,000 words in the vocabulary of a typical college graduate, or perhaps the total number of human faces a gregarious person learns to recognize over a lifetime.

When a chessmaster is confronted with a chessboard on which the pieces are arrayed in a "reasonable" way, he can store this information in short-term memory in a half dozen or fewer "chunks" -- familiar configurations. The ordinary player, or the chessmaster confronted with a randomly arrayed chessboard, must store the information piece by piece, hence can hold the positions of only about a half dozen pieces in short-term memory.

The skill that the expert acquires does not simply consist in being able to recognize familiar stimuli or configurations of stimuli. As deGroot showed, the recognition of perceptual features on the chess board reminds the grandmaster of moves that are potentially good when those features are present. Indeed, we should expect that the expert's knowledge for pattern recognition is integrated with strategic knowledge so that the patterns an expert has learned to recognize are those that are relevant to choices of moves and plans in game situations.

The importance of game strategy in perception and representation of complex patterns was shown in an experiment by Eisenstadt and Kareev (1975). Go and Gomoku are games with entirely different rules, but are played on the same board, and with the same kinds of pieces. Two groups of subjects, who knew how to play both games, were shown the same patterns of stones on boards, but in the one case were told that the patterns were from a game of Go, in the other case from a game of Gomoku. They were subsequently asked to recall the patterns. The subjects in the first condition had better recollection of the pieces that were critical to

selecting the correct move in the Go position, while the subjects in the second condition tended to recall those that were critical to selecting the move in the Gomoku position. Thus, in the face of a complex stimulus situation, attention to a particular task will determine the sequence in which information is extracted from the stimulus, and the patterns in which it will be organized.

Specific knowledge structures that integrate strategic knowledge and knowledge for recognizing patterns have been studied by Wilkins (1980), in a model of choosing moves in chess, and by Reitman and Wilcox (1978), in a model of playing Go.

Wilkins' (1980) model represents board positions by recognizing concepts, such as Attack and Safe, based on relations among pieces. The model uses schemata that correspond to the concepts in proposing and evaluating plans. In formulating a plan, a concept such as Safe or Defend-Threat can be set as a goal; the schema for each concept includes conditions that are required to satisfy the goal. The model's strategy of using proposed plans to guide its search restricts the set of moves it considers, enabling relatively thorough evaluations. The model is successful in solving problems of choosing moves in middle game positions that are difficult enough to be used in a standard chess textbook.

Reitman and Wilcox's (1978) model simulates representation of board positions and changes of board positions in Go. The model forms a multilivel representation with low-level units such as strings and chains of stones, and higher-level units called groups and fields involving collections of points and their surrounding stones. The representations include features that are relevant to Go tactics, such as the stability of a group of stones. Perceptual activity is organized according to several structures including lenses, which monitor changes on the board relevant to relations between groups of stones, and webs, which monitor changes on radii and circumferences around groups. The model's capabilities for representation, combined with some relatively low-level processes for selecting moves, is similar to human player with experience of playing 40 or 50 games.

The ability of experts to recognize complex patterns of information related to a highly integrated structure of actions has been found in other domains in which expertise has been analyzed. We have discussed the importance of knowledge for representing problems in physics in Section II.D. Analyses that we discuss in Section IV.D of knowledge for expertise in medical diagnosis and electronic troubleshooting have led to similar conclusions. A conjecture that is reasonable on present evidence is that high levels of expertise generally require a reportory of tens of thousands of perceptual "chunks" relevant in the domain. In domains where the minimal time required to become a world-class master has been measured, the estimate turns out consistently to be about a decade (Hayes, 1971; we discuss this finding for musical composition in Section III.D.1).

III.C. Construction Tasks and Other Insight Problems.

Much research attention has been given to problems in which some physical device or arrangement is required, often to satisfy a functional criterion. An example is Duncker's (1935/1945) famous "tumor" problem. A patient has a stomach tumor, which is to be destroyed by radiation without damaging the surrounding healthy tissue. How is it to be done?

The source of difficulty in construction problems is rather different from the problems discussed in Sections III.A and III.B, where difficulty is caused primarily by the large number of possible solutions. The tumor problem, and other "insight" problems, are difficult primarily because most of the candidate solutions that people think of are ruled out by the constraints of the problem. In the tumor problem, for example, one cannot simply direct the rays to the tumor, because that would destroy all the tissue along their path; one cannot open a path to the tumor because the surgical procedures would cause intolerable damage, and so on. The "textbook" solution of the tumor problem is that the tumor should be irradiated from many different angles, and hence via many different paths through the surrounding tissue. Then a large quantity of radiation can be concentrated on the tumor while each path of surrounding tissue is subjected to only a small fraction of that amount.

Solving the tumor problems and similar insight problems often depends on finding a way to represent the problem so that the solution becomes "obvious". Achievement of such a representation, corresponding to a moment of insight, is a phenomenon of great interest, especially in relation to issues of cognitive organization in Gestalt psychology. In problems such as cryptarithmic and anagrams, the problem space is easily constructed, and problem-solving activity consists of search in the set of possibilities that arise in that space. In contrast, in insight problems such as the tumor problem, the problem solver's initial representation usually provides an inadequate problem space, one in which a solution will not be found. Problem solving involves a construction of several problem spaces, with discovery of factors that make each of them inadequate, until a successful representation is found. Processes of problem representation, discussed in Section II.D, thus play a central role in solution of these problems of construction. The process can be characterized as a search, where the possibilities are alternative ways to represent the problem. However, the usefulness of such a characterization is limited unless the set of alternative representations can be specified more definitely than we are able to at present.

Duncker (1935/1945) emphasized the demand, the condition to be met by the problem solution, as the chief source of solution proposals. The initial proposals are not unmotivated, but they are faulty in not attending to all the conditions a solution must meet. False analogies may produce inadequate solutions because the analogy does not match, on crucial dimensions, the actual situation. At the same time, Duncker stressed that the proposals are not produced by simple association (page 3):

In short, it is evident that such proposals are anything but completely meaningless associations. Merely in the factual situation, they are wrecked on certain components of the situation not yet known or not yet considered by the subject.

Occasionally it is not so much the situation as the demand, whose distortion or simplification makes the proposal practically useless.

By constructing a taxonomy of correct and inadequate solutions to the tumor problem, Duncker showed how the solution generation process can be understood as a process of means-ends analysis. His taxonomy can be depicted in outline form:

- Treatment of tumor by rays without destroying healthy tissue
 - Avoid contact between rays and healthy tissue
 - Use free path to stomach
 - Use esophagus
 - Remove healthy tissue from path of rays
 - Insert a cannula
 - Insert protective wall between rays and tissue
 - Feed substance that protects
 - Displace tumor toward surface
 - By pressure
 - Desensitize the healthy tissue
 - Inject desensitizing chemical
 - Immunize by adaptation to weak rays
 - Lower intensity of rays through healthy tissue
 - Postpone full intensity until tumor is reached
 - Use weak intensity in periphery, strong near tumor
 - By use of lens

Duncker described the solution process as successive development or reformulation of the problem. Both working forward and working backward may contribute to the process. Seeing a stick may give a chimpanzee the clue to obtaining a banana that is out of reach. Alternatively, the banana's being out of reach may lead the chimpanzee to look for objects that could be used to reach it (cf. Kohler, 1929). Mistakes may also call attention to features of the problem situation that must be incorporated in the solution -- hence may lead to new solution attempts.

From the idea that problem solution depends on an appropriate formulation, it would be expected that hints could be used to make problems significantly easier. One experiment on the effect of hints used a problem of constructing a hatrack, invented by Maier (1945). Two sticks and a clamp were given, and the hatrack could be constructed by clamping the sticks together so they could be wedged between the floor and the ceiling. Subjects usually began by placing one stick on the floor, clamping the other stick to it vertically, or in an X or inverted V shape with one end of each stick resting on the floor. Neither of these structures is stable. If the experimenter said, "In the correct solution, the clamp is used as the coat hanger," solution was facilitated somewhat, mainly by reducing attempts with one stick lying on the floor. If the experimenter said, "In the correct solution the ceiling is part of the construction," solution was facilitated more strongly, with reduction both of attempts that have one stick on the floor and of attempts that have one end of each stick on the floor (Burke, Maier & Hoffman, 1966).

A potential source of problem solutions is analogy with similar problems. Gick and Holyoak (1980) gave Duncker's tumor problem to subjects some of whom had studied a story in which a fortress is taken by a converging attack. The subjects who were familiar with the military problem were more successful than control subjects in solving the tumor problem. An important factor was use of an instruction that the story might provide a useful hint for solving the problem. With the hint, most subjects found the convergence solution to the tumor problem, but without the hint only about one-half as many subjects found that solution, even though they had read the story and recalled it in a test.

In a subsequent study, Gick and Holyoak (1983) examined conditions favoring spontaneous use of an analogy. There was little effect of asking subjects to summarize the military story, rather than to recall it, nor did giving a verbal statement or a diagram showing the convergence principle notably increase subjects' use of the analogy. However, more solutions were given by subjects who read two stories involving convergence, summarized both of them, and discussed ways in which the stories were similar. Gick and Holyoak concluded that subjects acquired a schema with the idea of convergence represented in a general way, and that use of such a schema is more likely than use of a specific analogous problem. (Recall the use of a similar hypothesis in interpreting learning a subtraction procedure with understanding, based on an analogy with blocks, discussed in Section II.D.2.)

Duncker (1935/1945) also studied problems that required subjects to construct or assemble some item out of potential components that were provided. He showed that the problems could be made difficult by presenting one of the components in such a way that it was conceptually "unavailable" for its required function. In one problem, for example, the building materials were a candle, matches, and a box full of thumbtacks. The task was to mount the candle on a wall so that it could burn without dripping wax on the floor. The problem could be solved by thumbtacking the box to the wall, then mounting the candle in it.

This problem is sufficiently difficult that fewer than half the subjects in one experiment were able to solve it in 20 minutes (Adamson, 1952). When the problem was presented to another group of subjects with the thumbtacks lying on a table, and the box empty, 86 per cent solved it in less than 20 minutes. The phenomenon underlying this finding has been labeled "functional fixity." When an object is performing one function, or has recently been used to perform some function, subjects are less likely to recognize its potential use for another function.

Birch & Rabinowitz (1951) demonstrated a similar phenomenon, using another problem studied originally by Maier (1931). Subjects were introduced to a room where two strings were hanging from a ceiling, too far apart to be reached simultaneously. The task was to tie them together. This could be accomplished if a heavy object were tied to one string and the string swung as a pendulum. The subject could reach this string as it swung toward him or her while he or she was grasping the other string. Two objects, an electric switch and a relay, were available for constructing the pendulum. The subjects had used either the switch or the relay (but not both) in a previous task. Of ten subjects who had used the relay previously, all used the switch to construct the pendulum; of nine who had

used the switch, seven used the relay to construct the pendulum. Of six subjects who had used neither object previously, three used the switch and three the relay to construct the pendulum.

Several findings support a hypothesis that functional fixity results from a decrease in the likelihood of noticing certain critical features of objects in the situation, such as the flatness of a box (in use as a container), or the heaviness of a switch (after use in a circuit), or the mobility of a string. The mechanisms that produce decreased noticing of features in functional fixity may be quite different in different situations, involving restrictive hypotheses about general classes of solutions in some cases, and simple competition between feature-recognition processes in others.

Some of the findings that support this explanation involve demonstrations that solution of problems can be strongly influenced by quite low-level perceptual factors. For example, in the pendulum task, the idea of making the one string swing to make it reachable by someone holding the other does not occur readily to most subjects, even in the presence of one or more heavy objects. Maier (1931) showed that this idea occurred immediately to many subjects who had not previously thought of it when the experimenter casually brushed against the string and set it swinging. Glucksberg and Weisberg (1960) presented pictures of the materials available for use in solving Duncker's candle problem, and found that solutions were markedly increased when the label "Box" was included in the picture. A process of noticing features of objects that can be related to the problem goal (Duncker's "suggestions from below") probably plays a significant role in solution of construction problems, as Weisberg and Suls (1973) concluded in their theoretical analysis of solution processes for the candle problem. Results consistent with that idea were obtained by Magone (1977), who found that subjects produced a greater variety of solutions in Maier's two-string problem if they were initially prompted to consider features of objects than if they were initially prompted to find a solution of a specified kind, such as extension of one of the strings or causing a string to swing back and forth.

The Einstellung effect discussed in section II.3.2 is similar in character to functional fixity; both effects result from the influence of previous experience upon the availability of alternative solution steps for problems. Furthermore, the processes responsible for the two effects probably are analogous in a subtle but significant way. Both involve a condition in which a form of search is made less likely than it would be normally. In the case of Einstellung, previous use of one solution path suppresses a search for problem-solving operators. In the case of functional fixity, a search for features of objects that could be useful in the problem is suppressed.

Another "insight" problem that has been studied is the nine-dot problem. A three-by-three matrix of dots is given, and the task is to connect all the dots with four straight lines without any retracing. The problem is difficult; most subjects do not think of drawing lines outside the space defined by the matrix of dots, as is required for the solution. The difficulty is apparently another instance of a restricted domain of search, but the obvious hypothesis of a restriction based on the spatial arrangement is not supported by data. Weisberg and Alba (1981) instructed

subjects that they should draw lines outside the square of dots, but that had little effect. However, when they gave an easier problem requiring drawing lines beyond the region that contained dots, solution of the nine-dot problem was facilitated. A reasonable interpretation is that the easier problem led the subjects to consider problem-solving operators that were not in the problem space of subjects who did not solve the simpler problem first. This finding involves the same principles as the finding of Katona (1940; Section III. B.1) that a heuristic for choosing operators is more effective than a test applicable to the results of operators.

III.D. More Complex Tasks of Composition and Design.

In this section we discuss analyses of several complex tasks of composition and design. We begin with composition of written essays and music, tasks requiring methods for handling multiple constraints. Then we discuss two tasks of answering examination questions and designing policies in which the problem solvers' recognition and knowledge of constraints play an important role in successful performance. Third, we discuss design of procedures, in which the materials are actions that can be performed and the task is to use those actions to construct a procedure, such as a computer program, that satisfies a criterion.

III.D.1. Problems of Composition. We discuss an analysis of composing written essays, developed using thinking-aloud protocols by Flower and Hayes (1980), and an instructional study of writing by Bereiter and Scardamalia (1982). Then we briefly describe a study of composition in music by Reitan (1965), and data that show the need for extensive training in a domain as a condition for creative accomplishments (Hayes, 1981).

Flower and Hayes (1980) have studied the task of writing an essay, a task in which constraints play a major role. They noted that successful writing requires simultaneous compliance with a large number of constraints, operating at different levels. One requirement is selection and organization of ideas from the writer's knowledge into a coherent network of concepts and information for inclusion in the essay. Another set of constraints involves the linguistic and discourse conventions of written language. A third set of constraints is rhetorical, involving the need to arrange the essay so it accomplishes the writer's purpose for the intended audience.

Using protocols obtained from subjects working on writing tasks, Hayes and Flower (1980) found three general processes: planning, translating, and reviewing. These three processes allow the writer's attention to a subset of the constraints at any time. In planning, information is generated from the problem solver's memory relevant to the topic, and decisions are made about what to include. In translating, text is produced using information that has been retrieved, consistent with a writing plan that has been formed. In reviewing, the generated text is evaluated and revised in accord with rhetorical purposes and constraints of text structure, as well as more detailed linguistic concerns such as correct grammar. Hayes and Flower found that writing involves a mixture of these processes, and postulated that the writing process includes a monitor that determines the sequence of subprocesses, depending on the nature of difficulties that arise.

To write successfully, an individual must understand the constraints that apply at various levels to the text, must have effective methods for generating or revising text to conform to those constraints, and must actively engage in evaluation with respect to the constraints. In studies of young writers, Bereiter and Scardamalia (1982) noted that inattention to constraints, especially global rhetorical concerns, characterizes the writing of many children. When they revise text that they have produced, most children attend exclusively to low-level constraints, usually changing only single words or small phrases, rather than attempting to improve more significant general features of their essays. Bereiter and Scardamalia hypothesized that the difficulty lies in the process of evaluating the text, rather than in a lack of understanding of rhetorical goals or a lack of effective means for producing improved text. They gave students a set of cue cards with evaluative comments, such as "I need another example here," "The reader won't be convinced by this," "Even I seem to be confused here," and "This is a good sentence." The children's task was to choose a card that seemed appropriate for each sentence in their texts, and to make appropriate changes. The technique was effective, consistent with the idea that children have difficulty managing the process of evaluating their texts and applying global constraints, rather than lacking knowledge of the constraints and methods for complying with them.

Multiple interacting constraints also characterize composition of music, as was shown in an analysis by Reitman (1965) based on a protocol obtained from a professional composer as he wrote a fugue. Reitman noted that schematic structures that he called transformational formulas played an important role; these included knowledge of the main components of the musical form being composed (exposition, development, and conclusion) as well as subcomponents of those units (exposition \Rightarrow thematic material + countermaterial; thematic material \Rightarrow motive + development, etc.). He found that much problem-solving activity was concerned with constraints. Some constraints were generated by properties of the instrument (piano) chosen for the piece, requiring musical material suited to the instrument. Other constraints were produced by material already included in the piece, such as a requirement that countermaterial should be compatible with thematic material, but sufficiently different to provide interest. The composer characterized patterns that he developed as conventions, producing melodic, rhythmic, and instrumental properties that were then "used to carry on the movement of the music" (Reitman, 1965, p. 169), with variations introduced to maintain interest.

A substantial knowledge base is required for solving problems of composition, and an important question is how much experience and training an individual needs to make substantial creative contributions to a field such as musical composition. Using data from biographies and a standard catalogue of recordings, Hayes (1981) determined the time between a composer's beginning serious musical training and the first composition that had five independent recordings in the catalogue. In almost every case, at least ten years of virtually full-time training occurred before a composer produced a work of such high quality that it is common in the recorded repertoire.

III.D.2. Recognition and Knowledge of Constraints. In problems that require satisfaction of constraints, a problem solver must recognize the constraints in order to perform successfully. In Section III.A.1 we discussed Newell and Simon's (1972) finding that individual differences in cryptarithmic depended on inclusion in their problem spaces of significant constraints, such as odd-even parity. Now we discuss two more studies that have investigated this factor. In a study of performance on examination questions, Bloom and Broder (1950) found that some students performed poorly because they failed to recognize constraints stated in the questions. In a study of problems involving administrative policy, Voss, Greene, Post, & Penner (1983) found that novice problem solvers gave inadequate formulations of problems because they lacked knowledge of significant factors in the problem situation.

In comprehensive college examination questions studied by Bloom and Broder (1950), students often were required to make inferences or deal with information presented in an unusual form. For students who performed poorly, a significant factor was the students' inattention to constraints in the statements of some questions. For example, a question might ask a student to choose the best explanation of a situation, but the student would ignore the relation of alternative answers to the situation and pick the one that seemed most nearly true in itself. For these students, the activity of problem solving occurred in a problem space that lacked some of the information that was required for good performance. Bloom and Broder developed an instructional method in which students compared their own problem-solving process, recorded in a thinking-aloud protocol, with the process of another student whose performance was more successful. This training was effective for many students, teaching them to attend more carefully to constraints in questions as well as other helpful problem-solving strategies, such as increased attempts to infer plausible answers from information they could retrieve from memory.

Voss et al. (19xx) obtained thinking-aloud protocols on problems involving design of an administrative policy. For example, problem solvers were asked to develop a policy for improving agricultural productivity in a region of the Soviet Union. Subjects with different amounts of knowledge about Soviet government and history worked on the problem, including students in an introductory course in Soviet politics, experts in political science, some who specialized in the Soviet Union and others with other specialties, and experts in another field altogether, chemistry. The solution process of experts was primarily one of formulating the problem, with a long initial period of considering historical and political factors and successive reformulations based on evaluations of proposed solutions against known constraints. The inexpert student subjects, with much less knowledge than the experts, gave problem formulations that lacked inclusion of important constraints. Experts in chemistry worked more systematically than the political science students, sometimes using general knowledge about administrative systems to provide useful conjectures, but also lacked the rich formulations that characterized the problem solvers with specialized knowledge.

III.D.3. Design of Procedures. To complete this section, we discuss tasks in which the materials are a set of actions that can be performed, and the problem solver constructs a procedure with these components. In a simple example, a list of errands was shown to subjects, along with map,

and the subjects constructed a schedule for performing as many of the errands as possible during a day (Hayes-Roth & Hayes-Roth, 1978). More complex examples are computer programming and software design, and design of experimental procedures in microbiology. We have mentioned the similarity of these problems to problems of transformation, discussed in Section II, especially when planning is used for constructing a sequence of actions to reach the problem goal. The knowledge structures of experienced problem solvers that have been analyzed in domains of software design and design of experimental procedures are very similar to the domain-specific knowledge used in planning, for example in the domain of geometry discussed in Section II.B.1.

Hayes-Roth and Hayes-Roth (1978) gave subjects a map of a fictitious town, showing the locations of several stores and other businesses. Subjects also were given a list of errands, such as buying fresh vegetables at the grocery, picking up medicine for a dog at the vet, and seeing a movie. Subjects were to plan a schedule that included as many of the errands as possible. The task presents some general constraints, mainly a limited total amount of time available. It also presents local constraints and interactions. For example, it is better to buy groceries later in the day, so they will still be fresh when the shopper returns home; and it is best to go to the movie at one of the times that the feature is starting. Interactions include proximity of shops, making it more efficient to place errands together in the sequence that involve shops that are near one another.

Hayes-Roth and Hayes-Roth simulated performance on their planning task with a model that contains several planning specialists and a blackboard control structure, a design similar to one used earlier in a speech understanding system called Hearsay (Reddy, Erman, Fennell, & Neely, 1973). The specialists are designed to make suggestions about different kinds of planning decisions. They all have access to inferences, suggestions, and other information, which is located in the system's blackboard. This system design supports a feature called opportunistic planning, which was found in the performance of human problem solvers. Opportunities arise in the form of conditions that make it easy to include an errand, such as the proximity of a store to a place that is already included in the plan, and an appropriate specialist can be activated by that condition.

In writing a computer program, one designs a procedure that performs a designated function. Studies of computer programmers and designers have revealed important characteristics of the knowledge required for solution of these design problems.

Soloway, Ehrlich, Bonar and Greenspan (1982) gave three problems, typical of elementary programming courses, to students in the first and second introductory courses in programming. They identified schematic cognitive structures that they called plans, needed for successful problem solving. The required schemata are quite basic, involving construction of iterative loops and use of variables. The schemata provide knowledge of requirements for performing significant program functions, such as interactions between processing and testing a variable within a loop and between the loop processing and initialization. Students who lacked adequate versions of these schemata made significant errors, for example, failing to recognize distinctions between different looping structures.

Experiments on memory for program texts have shown that experienced programmers can recall more successfully than beginners (Adelson, 1981; McKeithen, Reitman, Rneter, & Hirtle, 1981) a phenomenon like that found for memory of chess board positions (Section II.B.2) and other domains. The acquisition of plan schemata as hypothesized by Soloway et al. provides a natural explanation of this finding.

More advanced problems, involving software design, were studied by Polson, Atwood, Jeffries, & Turner (1981). A task in software design involves planning a complex program; actual writing of the program is performed separately. A problem studied by Polson et al. was design of a program for compiling an index for a text, given a set of key words to be included in the index. Problem solutions with thinking-aloud protocols were given by professional software designers and by students. The experts recognized functions that had to be included in the solution, such as defining a data structure for the text and searching the key word set for a word that matches each word encountered in the text. Polson et al. concluded that experts' knowledge includes general design schemata that enable decomposition of problems, progressively forming more well-defined subproblems, with knowledge of specific techniques for some subproblems that are encountered. These schemata provide another example of knowledge for action organized hierarchically like that developed by Sacerdoti (1977; Section II.B).

Problems of experimental design have been studied in the domain of microbiology; two versions of a program called Molgen have been developed. One, by Stefik (1981), designs procedures for modifying the genetic structure of microorganisms. An important issue considered by Stefik was the handling of constraints that arise from interactions between components of a procedure. Molgen designs procedures in a top-down manner, with abstract plan schemata gradually made more specific. A method of constraint posting was developed in which requirements for one of the design components can be taken into account in the decisions made about other components.

The second version of Molgen, by Friedland (1979), designs analytic experiments such as determining the sequence of base molecules in a DNA string or finding a set of restriction sites on a molecule. This model uses schemata called skeletal plans that incorporate information about experimental procedures that are used to develop specific experimental plans through a process of filling in details, based on the specific problem requirements.

IV. Induction

In this section we discuss inductive problem solving. In a problem of induction, some material is presented and the problem solver tries to find a general principle or structure that is consistent with the material. Important examples include scientific induction, including situations in which the material is a set of numerical data and the task is to induce a formula or a molecular structure, language acquisition, where the material is a set of sentences and the task is to induce the rules of grammar for the language, and diagnosis, where the material is a set of symptoms and the task is to induce the cause of the symptoms. Analogy problems and problems of extrapolating sequences are inductive tasks that are widely used in intelligence tests. The task of inducing a rule for classifying stimuli into categories has been used in a large and significant body of experimental study.

An induction problem presents a dual problem space that includes a space of stimuli or data and a space of possible structures, such as rules, principles, or patterns of relations (cf. Simon & Lea, 1974). The task can be conceptualized as a search in the space of structures. The problem is solved by finding a structure that satisfies a criterion of agreement with the stimuli or data. An experimental subject can be tested by requiring use of the structure for stimuli that have not yet been shown. When the task is to induce a rule for classifying stimuli, new stimuli may be presented to test whether the subject can classify them correctly. When the task is to induce a pattern in a sequence, the subject may be required to extend the sequence by producing additional elements that fit the same pattern as those that are given.

Solving an induction problem can proceed in two ways, and in most tasks a combination of the methods is used. The first is a top-down method. It involves generating hypotheses about the structure and evaluating them with information about the stimulus instances. Second, there is a bottom-up method that involves storing information about the individual stimuli and making judgments about new stimuli on the basis of similarity or analogy to the stored information. To perform the top-down method, the problem solver requires a procedure for generating or selecting hypotheses, a procedure for evaluating hypotheses, and then a way of using the hypothesis generator to modify or replace hypotheses that are found to be incorrect. To use the bottom-up method, the problem solver needs a method of extrapolating from stored information, either by judging similarity of new stimuli to stimuli stored in memory or by forming analogical correspondences with stored information.

Induction involves a form of understanding, where a representation is found that provides an integrated structure for diverse stimuli. This general feature also characterizes processes of representing problems (e.g., textbook physics problems) discussed in Section II.D. There, the space of stimuli is the information in the problem situation -- often, a problem text or instructions -- and the space of structures is a set of possible representations that can be constructed. To be successful, a problem representation must provide the information needed to achieve the problem goal. Thus, in representing transformation problems, the inductive search is constrained by the requirements of problem-solving operators that are available. Often in problems of induction, such constraints are not

present: one does not have to do anything with the pattern that is found in the information. However, in some inductive problems, such as medical diagnosis, there are strong constraints related to available operators. The goal is to restore the ailing person to proper functioning, and the effort to induce a cause serves the goal of determining an effective remedy.

In some task domains, the possible structures are represented explicitly as formulas. Examples include induction of quantitative formulas from numerical data in physics, or induction of the molecular structure of a chemical compound. Patterns induced in letter-sequence problems also consist of explicit formula-like rules. These tasks share important properties with problems of design and arrangement, discussed in Section III. The goals of these induction tasks can be considered as design of a formula that agrees with the data. Solution of design problems generally requires use of strong constraints to limit the space of possibilities for search, and this important property is also found in tasks that involve induction of formulas.

Our discussion of inductive problem solving will be in four subsections. In Section IV.A, we discuss induction of categorical concepts, including induction of rules for classifying stimuli and categorical concepts in the form of prototypes. In Section IV.B we discuss induction of more complex concepts involving sequential stimuli: patterns in sequences of letters and the grammatical rules of a language. Section IV.C discusses induction of relational structure including analogy test items and induction of regularities and structures in science and mathematics. Finally, Section IV.D discusses diagnostic problem solving, including medical diagnosis and electronic trouble-shooting.

IV.A. Categorical Concepts

Of the various inductive tasks that have been studied, by far the greatest attention has been given to induction of categorical concepts. This is partly in recognition of their practical importance. Our human capability of organizing experience using conceptual categories undoubtedly contributes much to making our cognitive lives manageable.

In an experiment on concept induction, the experimenter constructs a set of stimuli (e.g., diagrams with figures that vary in shape, size, color, and other attributes), and decides on a rule for classifying the stimuli (e.g., "the red circles are positive, all other stimuli are negative"). The subject is given information about several individual stimuli -- that is, is told whether each stimulus is positive or negative. The subject's task is to induce the rule of classification. Usually, the experimenter tests whether the subject has induced the concept by presenting new stimuli to determine whether the subject can classify them correctly.

In an early discussion, Woodworth (1938) distinguished between processes of concept induction involving bottom-up and top-down methods. In a bottom-up process, knowledge of the concept is analogous to a composite photograph, consisting of an impression summed over the various stimuli in the category with the common features emphasized and the variable characteristics "washed out." In a top-down process, the

problem-solver actively constructs hypotheses about features that define the concept and tests these hypotheses with additional information about examples.

We discuss research on concept induction in three subsections. The first two discuss studies of top-down processes of inducing concepts that are defined by two or more stimulus features and then of concepts defined by a single feature. The third subsection discusses studies of bottom-up processes of inducing concepts.

Figure 11 here

IV.A.1. Multifeature Concepts. When two or more stimulus features are combined to form a categorical concept, they are combined in some logical formula, such as "A and B," or "If A then B." A stimulus is a positive example of the concept if the formula is true as a description of the stimulus. Consider the set of stimuli shown in Figure 11. The concept "Green and Circle" specifies the stimuli in the second column from the left. The concept "Green or Circle" specifies the stimuli that are in columns 1, 2, 3, 5, and 8.

Consider the requirements for performance of this task, assuming that it is done in a top-down, hypothesis-testing manner. First, the stimulus features must be discriminated; the problem solver must have processes for recognition of the features that are used to define concepts. Second, there must be a process for hypothesis formation, which constructs candidate hypotheses to be considered. Third, a process of hypothesis evaluation is needed to test the hypotheses that have been formed. Finally, a process for hypothesis modification is required to use the results of the tests to eliminate incorrect hypothesis, to change existing hypotheses or form new ones.

A landmark study of multifeature concept induction was conducted by Bruner, Goodnow, and Austin (1956). Bruner et al. observed subjects working on concept induction problems, including verbal reports about their hypotheses. The results we discuss here were from experiments in which subjects were instructed that concepts were conjunctions of features, but had to induce how many features were relevant and what the features were. We consider two experiments in which the problem was to induce a concept defined as the conjunction of two features.

In one experiment, subjects were required to solve two problems with the array of Figure 11 shown, and a third problem of the same kind "from memory," that is, with the stimuli not available. In each case, the problem began with the experimenter providing a positive instance, a stimulus that was a member of the concept category. Then the subject could choose any stimulus in the display and ask whether it was a positive or negative instance of the concept. The subject could offer a hypothesis after the choice of a card, but this was not required. The subject continued choosing cards and receiving information until the correct concept was induced.

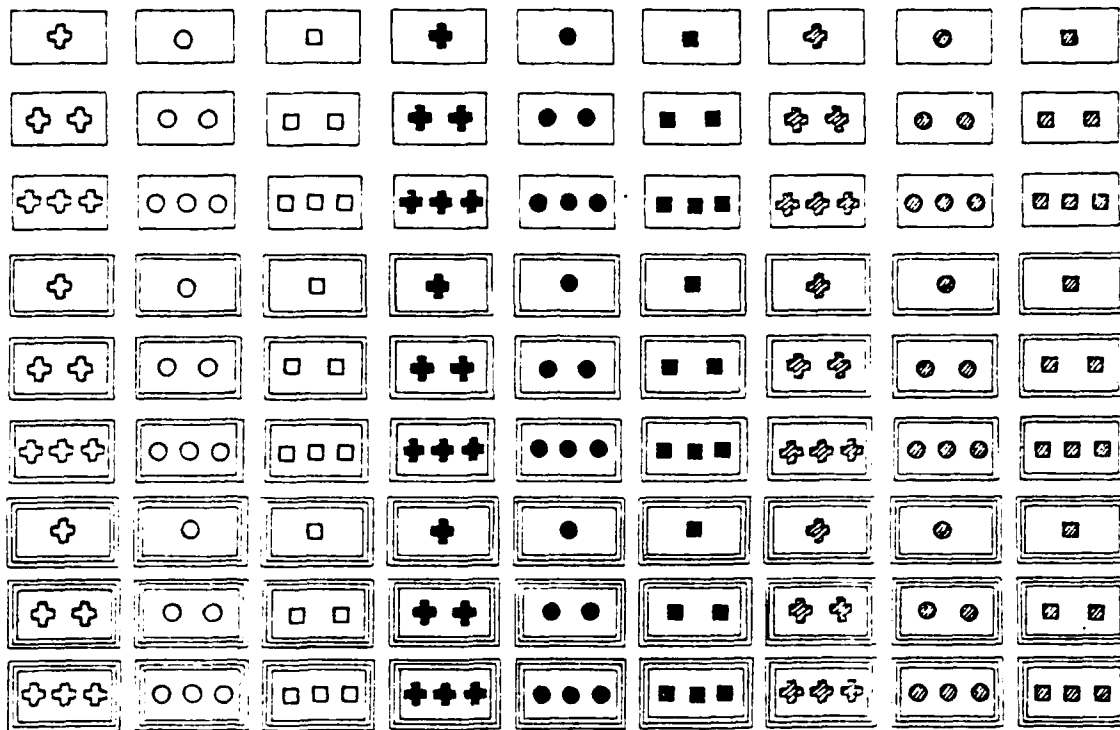


Figure 11. An array of instances comprising combinations of four attributes, each exhibiting three values. Open figures are in green, striped figures in red, solid figures in black (from Bruner, Goodnow, & Austin, 1955).

Bruner et al.'s results included characterizations of a variety of strategies used by subjects in selecting stimuli. Strategies of one kind, called focussing strategies, involve finding a positive instance of the concept, then determining which of its features are relevant. For example, suppose the concept was "Red and Circle." The subject might be told that the stimulus with three red circles and two borders is a positive instance. The subject then could choose a stimulus differing from that focus stimulus in the number of circles, say, two red circles with two borders. This would be a positive instance, and the subject would infer that the number of figures is not a relevant attribute. Then the subject might vary the color of the figures, choosing the stimulus with three green circles and two borders. This would be a negative instance, and the subject would infer that the color of the figures is relevant, that is, that "Red" is part of the definition of the concept. With further choices and information, the concept's definition would be inferred.

Other strategies, called scanning strategies, involve consideration of specific hypotheses and use of information to narrow down the set of possible hypotheses. For example, a subject might consider as distinct possibilities the hypotheses "Three figures," "Red," "Three and Red," "Circle," "Three Circles," and "Red Circles." Then finding that a stimulus with two red circles and two borders is a positive instance, all the hypotheses with the property "Three" could be eliminated. Use of a scanning strategy places severe demands on memory. It is impossible to consider all of the possible hypotheses simultaneously (there are 255 of them), but it is desirable to consider as many as one can, since information can only be used to evaluate hypotheses in the sample being considered.

The focussing strategies and the scanning strategies differ primarily in the processes they use in formation of hypotheses. In the focussing strategies, information about instances is used to constrain hypothesis formation. Tests are performed to see whether an attribute is relevant, and when the attribute is eliminated, no hypothesis using it will be formed. If the focussing strategy is used successfully, all but the correct attributes can be eliminated, and the correct hypothesis can be formed directly. In the scanning strategies, less use of problem information is used in forming hypotheses, and hypotheses that are in the sample are tested directly with information about instances. The use of information in evaluating hypotheses is somewhat more direct in the scanning strategies, but there is a consequently greater requirement for record-keeping in memory regarding a large set of hypotheses.

There were 12 subjects in Bruner et al.'s experiment, and their performance was used to classify them as focussers and scanners. A subject was classified as a focusser if the majority of his or her choices differed in just one attribute from features of the focus stimulus that had been found relevant or were as yet untested, or involved specific variations of this selection process, including redundant tests or attempts to test more than one attribute with a stimulus. Seven subjects were classified as focussers and the rest were treated as scanners. The focussing strategy was advantageous for the subjects who used it. They required about one-half as many choices as the scanners to solve a problem with the stimulus array present (medians of five and ten choices, respectively, for focussers and scanners). Further, the scanners had noticeably greater

difficulty solving a problem "in their heads" than they did when the stimuli were present (median of 13 choices), except for one scanner who discovered the focussing strategy in working on the third problem. The focussers' performance without stimuli present did not differ from their second problem with the stimuli.

Bruner et al. conducted two experiments to investigate situational factors that influenced subjects' choices of strategies. One experiment compared the effect of an orderly arrangement of stimuli with the same stimuli presented haphazardly. The stimuli used abstract forms, differing on six dimensions with two values on each dimension. With the 64 stimuli arranged systematically, similar to the arrangement in Figure 11, almost all subjects used focussing strategies. When stimuli were not arranged systematically, subjects typically used scanning strategies. Another condition in which there was a tendency to use scanning strategies was when concrete stimuli were used, such as drawings of persons varying in sex, size, and clothing.

An analysis by Hunt (1962; Hunt, Marin & Stone, 1966) provided a hypothesis about how to represent categorical concepts in cognitive structure. Hunt proposed that knowledge of a categorical concept is a cognitive procedure for deciding whether a stimulus is or is not a member of the category. The form of the procedure that Hunt investigated is a decision network, a structure of perceptual tests organized in a way that reflects the logical structure of the concept. (This same form was used by Feigenbaum (1963) for the Elementary Perceiver and Memorizer, used in simulations of rote verbal memorizing. Examples of such decision networks, for recognizing some concepts in geometry problems, are in Figures 5 and 6.) Experiments conducted by Trabasso, Rollins & Schaughnessy (1971) provided evidence that supports Hunt's characterization. Trabasso et al. measured latencies for categorical decisions about stimuli and obtained results that agreed with Hunt's model; longer times were required for decisions in which the model specifies a larger number of perceptual tests. A model that simulates acquisition of conjunctive concepts was developed by Williams (1971) using Hunt's representational hypothesis along with assumptions about limited short-term memory capacity and changes in the salience of dimensions.

An important aspect of the acquisition of complex concepts is induction of the logical relation between the stimulus features in the definition. This has been studied by Bourne and his associates in experiments in which subjects are informed of the features that the rules include. For example, a subject might be told that the rule includes "Red" and "Circle," but then would have to discover from examples whether the combination is conjunction, disjunction, conditional, or biconditional. When subjects are not experienced in this rule learning task, there are substantial differences in the difficulty of inducing the different kinds of rules, and these correspond to differences among the types of rules found in standard concept induction tasks (Haygood & Bourne, 1965). The rule that is learned most easily is the conjunctive rule.

One possible explanation for differences in difficulty is that the rules differ in familiarity to the subjects, with conjunction being the most familiar way to combine features. This would lead to a bias in the process of forming hypotheses, with the less familiar forms of hypothesis

generated later, if at all, and consequent delays in problem solutions. Evidence supporting such an interpretation was obtained by Bourne (1970), who found that differences among the rule forms decreased when subjects were given a series of rule induction problems. A more specific hypothesis, proposed by Bourne (1974), is that with experience, subjects acquire a strategy for representing information about stimuli in terms of truth-table values regarding the features known to be relevant. For example, if "Red" and "Circle" are the features, then a red circle has the value T-T (true on both attributes), a green circle has the value F-T, and so on. This is an efficient representation for solving concept-induction problems, because each of the alternative rule forms corresponds to a distinctive subset of truth-table values. A conjunctive rule is satisfied only by T-T; a disjunctive rule is satisfied by T-F, F-T, and T-T; a conditional rule is satisfied by T-T, F-T, and F-F; and a biconditional rule is satisfied by T-T and F-F. The truth-table hypothesis is supported by a finding by Dodd, Kinsman, Klipp and Bourne (1971) that training on a task of sorting stimuli into the four categories of the truth table facilitated subsequent performance on rule induction problems.

IV.A.2. Single-Feature Concepts. We now consider induction of conceptual rules consisting of single features, such as "all the red pictures," or "the circles." The task of inducing such a concept is simpler, of course, than inducing a multifeature concept.

Evidence for Top-Down Induction. Single-feature concept induction has been studied extensively by H. H. Kendler, T. S. Kendler, and their associates. One question addressed in their experiments is whether concepts are acquired in the form of a verbalized rule or in the form of an aggregation of individual stimulus-response connections. It is likely that a verbalized rule would result from a top-down hypothesis-testing process of induction, and an aggregation of stimulus-response connections would probably result from a bottom-up process.

Evidence has been obtained in experiments in which the conceptual category is changed without informing the subject. A subject is given an initial concept-induction problem involving a single stimulus feature (e.g., "respond positively to red stimuli"). After the subject meets a criterion of correct responding the rule is changed, either changing the positive value of the same attribute (e.g., from red to green), called a reversal shift, or changing to a different attribute (e.g., from red color to large size), a nonreversal shift. It was found that adult human subjects, and kindergarten children who solved the initial problem quickly, adjusted more easily to the reversal than to the nonreversal shift (Buss, 1953; Kendler & D'Amato, 1955; Kendler & Kendler, 1959), while rats and slower-learning kindergarten children adjusted more quickly to the nonreversal shift (Kelleher, 1956). An interpretation is that adults and school-aged children use a hypothesis such as "it depends on color," which does not have to be changed to adjust to the reversal shift, while nonhuman subjects and preschool children learn specific stimulus-response associations, for which the reversal shift requires a greater change. In a later study, Erickson (1971) found that college-student subjects adjusted more rapidly to nonreversal shifts if they had been carefully instructed about the nature of the concept induction task, suggesting that when subjects have more complete information about the task, they tend to remove stimulus attributes from consideration when their hypotheses are

disconfirmed.

Further evidence that adult human performance in concept induction is based on definite hypotheses was obtained by Levine (1963) who showed that on a series of test trials with no feedback given, nearly all of the sequences of responses given by college-student subjects were consistent with a systematic hypothesis about the conceptual rule.

Processes of Sampling Hypotheses. The processes of forming and evaluating hypotheses in single-feature concept induction are quite straightforward. Any stimulus feature that is noticed can be the basis of a rule, and a rule that links a feature with a response is confirmed or negated directly by information about the category of any example. Because the hypotheses are simple, and there are many possible hypotheses, it is efficient for subjects to consider samples of hypotheses rather than a single hypothesis at a time. A sample of a few hypotheses is considered, and on each trial the subject can eliminate hypotheses that are inconsistent with the information given about that trial's stimulus. If the sample includes the correct hypothesis, the process of elimination can narrow the sample down to that hypothesis, which solves the problem. If the sample does not include the correct hypothesis, eventually all the hypotheses in the sample will be eliminated and the subject will have to generate another sample. Note that this method is similar to the strategies that Bruner et al. (1956) called scanning. Like the scanning strategies, the strategy of testing samples of hypotheses is demanding on memory.

Theoretical discussions have included many proposals about processes of choosing hypotheses to consider, eliminating hypotheses based on stimulus information, and remembering previously eliminated hypotheses. Several of the proposals have been discussed by Gregg and Simon (1967) and by Millward and Wickens (1974).

Results obtained by Wickens and Millward (1971) provide support for an assumption that experienced subjects remember stimulus attributes that they have eliminated. According to Wickens and Millward's model, if the sample of hypotheses is exhausted, the attributes of eliminated hypotheses are stored in memory. Memory limitations apply both to the size of the sample that can be considered and to the number of previously eliminated attributes that can be remembered. In Wickens and Millward's experiment, subjects received extensive training in concept induction, solving many problems with the same set of stimuli, with different attributes used to define the concept in the successive problems. Performance improved sharply after the first problem or two, and stabilized within 10-20 problems. The model of attribute-elimination was supported by statistical data as well as by subjects' responses to a retrospective questionnaire. Differences in performance among the individual subjects can be explained by assuming that they all performed in accord with the model's assumptions, but that they differed in the sizes of the hypothesis samples that they considered and in their capacities for remembering previously eliminated hypotheses.

When performance of inexperienced subjects has been analyzed using stochastic models, the results have revealed a surprisingly simple structure of the problem-solving process. Restle (1962) investigated mathematical properties of a process in which a sample of hypotheses is considered by the subject, and on each trial a response is chosen using one of the hypotheses. In Restle's model it is assumed that subjects' processing of information differs depending on whether the response on a trial happens to be correct. After each correct response, hypotheses that are inconsistent with the information about that trial's stimulus are eliminated from the sample. After an error, the subject considers a new sample of hypotheses. A simple stochastic process results if it is assumed that sampling occurs with replacement. If this assumption is made, solution of the problem is an all-or-none event; the probability of solving the problem with no more errors after taking a sample is a constant, independent of the number of trials or errors that have occurred previously. This implication is counterintuitive. If we assume that the subject is sampling and testing hypotheses, it says that there is no accumulation of information over trials that makes sampling of the correct hypothesis more likely. The all-or-none property also is incompatible with almost any assumption of learning stimulus-response associations that are strengthened gradually over trials, as well as the summative or "composite photograph" process that Woodworth (1938) discussed.

The counterintuitive all-or-none property of Restle's model received strong empirical support in experiments by Bower and Trabasso (1964). Their experiments with college-student subjects included conditions in which the categorical rule was changed before the subject solved the problem, either using a reversal shift or a nonreversal shift. The assumption of resampling with replacement after errors predicts that shifts prior to solution should not delay the solution of the problem, and this surprising result was obtained.

Computer simulation models of the concept induction task, using different hypothesis generating strategies, have been proposed by Gregg and Simon (1967). They showed that these process models can be aggregated (approximately) into simple stochastic models like Restle's (1962), providing an information-processing explanation for the simple statistical regularities implied by the stochastic models and found in Bower and Trabasso's (1964) data. Gregg and Simon found that a range of different models is required to account for the set of experiments reported by Bower and Trabasso. According to these models, the nature of sampling depends primarily upon how much information the subjects can retain about the classification of previous instances, and about which hypotheses had already been refuted by the evidence. In general, the process models that fit the data best were those that implied very severe restrictions on short-term memory for previous instances and their classification. Given this restriction on memory, the models are consistent with the all-or-none property -- that is, the expected number of trials to solve the problem is independent of the time the subject has already spent on it.

IV.A.3. Bottom-Up Induction of Concepts. In addition to inducing categorical concepts in a top-down, hypothesis-based manner, induction also can be a bottom-up process, involving gradual emergence of the concept from the features of individual stimuli. This idea has received less attention in psychological research, but it has not been totally missing from the

discussion.

Hull (1920) conducted a study of learning in which the materials were pseudo-Chinese ideograms paired with nonsense syllables. The stimuli paired with the same response syllable from list to list all shared a stimulus component, a radical that was part of each of the stimuli. Hull's subjects showed positive transfer on the later lists in the experiment, indicating that they had induced the concepts to some extent. However, they typically were not aware of the feature or features that were shared, indicating that they were not actively testing hypotheses about the categorical rules. It seems likely that the subjects stored information about the individual stimulus-response pairs, and gradually built up impressions that included the shared components.

A result similar to Hull's was obtained by Reber (1967), who studied induction of rules for an artificial language. Reber constructed sequences of letters using a set of grammatical rules: for example, "Start with a T or a V," "After an initial T, use a P or another T," and "After a V that is not at the beginning, use a P or end the sequence." The sequences, from six to eight letters long, were used in a learning task in which subjects were shown the sequences and had to recall them. Subjects working on the grammatical sequences learned faster than subjects who worked on a comparable set of random letter sequences. After learning a set of grammatical sequences, subjects were able to discriminate between new grammatical sequences and sequences that violated the grammar with more than 75% accuracy. Even so, subjects were not aware of the rules that were used to form the grammatical sequences, and showed little awareness of their shared features.

In recent research and discussion, Rosch (e.g., 1978) has argued persuasively that much of our conceptual knowledge is not organized on the basis of definite feature structures, like those used in most experiments on induction of categorical rules. First, Rosch, Mervis, Gray, Johnson, and Boyes-Brain (1976) argued, with empirical support, that concepts at different levels of generality are not equal in salience, but that there are basic categories whose members share large numbers of features that are not shared by members of other categories, including characteristic patterns by which we interact with them motorically. For example, "chair," "table," and "hammer" refer to basic categories, while their superordinates, "furniture" and "tool," and their subordinates, such as "picnic table" and "claw hammer," are less fundamental. Data supporting this distinction were obtained by Rosch et al., who gave subjects a series of 90 terms and asked them to write all the attributes that came to mind. Another group of subjects was given the same terms and asked to write descriptions of muscle movements that they would make in interacting with the objects. Many more attributes and movements were associated with basic-level terms than their superordinates, and very few attributes or movements beyond those for the basic terms were given for the superordinate terms.

Rosch (1973, 1975) also has argued that natural categories are represented as prototypes, rather than as sets of features. A prototype may be thought of as a kind of schema for representing a category, which is activated more readily by typical members than by atypical ones. For example, in the category of "birds," the prototype

AD-A138 889

PROBLEM SOLVING AND REASONING(U) PITTSBURGH UNIV PA
LEARNING RESEARCH AND DEVELOPMENT CENTER

22

J G GREENO ET AL. FEB 84 UPITT/LRDC/DNR/APS-14

UNCLASSIFIED N00014-79-C-0215

F/G 5/9

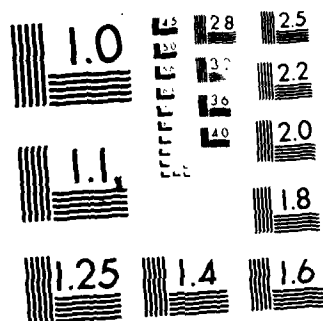
NL

END

DATE

FILED

DTIC



MICROCOPY RESOLUTION TEST CHART
NATIONAL BUREAU OF STANDARDS 1963-A

canaries are judged more typical than penguins or peacocks, and in the category of tools, hammers and saws are judged more typical than anvils or scissors. Rosch (1975) found that there is very strong agreement among subjects in ratings of typicality. Evidence that typicality influences cognitive processes has been obtained when subjects are asked to judge whether statements such as "A robin is a bird" or "An anvil is a tool" are true. In these experiments, judgments are made more quickly for the statements involving more typical examples (Rosch, 1973; Rips, Shoben & Smith, 1973).

Acquisition of prototypical concepts has been studied experimentally by Posner, Goldsmith and Welton (1967), Franks and Bransford (1971), and Reed (1972), among others. For these experiments, a set of stimuli is constructed by varying a single stimulus, the prototype. The stimuli may be geometric forms, patterns of dots, schematic faces, or other kinds of stimuli. The stimuli are shown to subjects, and then a recognition test is given. Subjects' confidence in recognition is a function of the similarity of stimuli to the prototype. When the prototype itself is shown, subjects respond positively with strong confidence, even if the prototype was not included in the set of stimuli they saw. Several investigators have shown that this performance can be explained by considering the frequencies with which various stimulus features occur during the learning trials; for example, the features of the prototype appear with high frequency, even if the prototype itself is not presented (Reitman & Bower, 1973; Neumann, 1974).

A model that simulates bottom-up acquisition of a prototypical concept has been formulated by Anderson, Kline & Beasley (1979), using general principles of learning in the context of a production-system model of performance. Anderson et al.'s system stores cognitive representations of the patterns seen in individual stimuli, and additional representations are stored by processes of generalization and discrimination. Representations are strengthened when they provide a basis for recognizing stimuli that are presented. Anderson et al.'s simulation accurately mimics subjects' performance on recognition tests, including false recognition of prototypes that have not been presented during learning.

A reasonable expectation is that many learning processes are not strictly top-down or bottom-up, but a combination of the two. Such a combination was analyzed by Greeno and Scandura (1966) and by Polson (1972) in studies of concept induction involving verbal items. In the experimental setup, like that used by Hull (1920), lists of paired associates are presented to be memorized, and in successive lists the same response term is paired with different stimuli that are related to one another. Greeno and Scandura found that transfer to individual items occurred in an all-or-none manner; different sets of items had differing proportions of items with no errors, but for items with any errors performance in the transfer conditions could not be distinguished from each other or from performance on control items. The finding of all-or-none transfer suggests a top-down conceptual process in which any individual item either is or is not recognized as a member of a definite category. Polson (1972) studied acquisition of the conceptual categories and found that this was not an all-or-none process. The findings were consistent with a hypothesis of a two-stage process. For some subjects, there is an initial stage of bottom-up learning, in which associations of responses

with patterns of features are stored, with transfer depending on features that are shared by similar items. In the initial phase, the subject may notice the shared features of members of a concept category by chance. Once the shared feature of a category is recognized, the second stage of learning occurs involving an active, top-down process in which the subject actively searches for features to use in classifying the stimuli.

It is likely that both the top-down and the bottom-up methods of learning about categories are available to human learners, and the question arises as to what circumstances make it more likely for one rather than the other to be used. Brooks (1978) compared a condition in which subjects were asked to learn names for individual stimuli with a condition in which subjects induced a rule for classifying stimuli. Explicit rule induction led to better knowledge of relevant features, reflected in better performance on classification of new stimuli, as would be expected from learning by top-down induction. Subjects who learned individual names gave superior performance in recognition of specific stimuli from the learning set, but also recognized new stimuli at an above-chance level, as would be expected from bottom-up acquisition of a concept involving a summation of instances.

IV.B. Sequential Concepts.

We now discuss two tasks involving induction of concepts that are more complex than those discussed in Section IV.A. In the tasks we discuss now materials are sequences of elements that are organized in patterns. The subject's task is to induce the patterns. First, we discuss the task of extrapolating sequences of letters, where a subject must identify patterns in the sequences that are presented and use the patterns to extend the sequences. The second task is induction of grammatical rules of a language from example sentences that are consistent with the grammar.

In these tasks, the problem space includes a set of stimuli and a space of possible structures, as in all induction problems. However, in comparison to the space of possible rules for classifying stimuli, the space of possible pattern descriptions for sequences and the space of possible grammatical rules are extremely large. To solve these problems, substantial reductions of the search spaces are required, and these are accomplished by constraints on the generation of hypotheses. In sequence extrapolation, a limited set of relations and sequence forms are considered. In the analysis of grammar induction that we discuss, hypotheses about the structures of sentences are constrained by the structures of situations that the sentences describe.

IV.B.1. Sequence Extrapolation. An example of a sequence extrapolation problem is the following: mabmbcmcdm __, where the task is to extend the sequence. In a model of sequence extrapolation formulated by Simon and Kotovsky (1963), a pattern is induced from basic relations between the letters in the problem string. The pattern is a kind of formula for producing the sequence; once discovered, the formula can be used to extend the sequence, as required.

For example, for the problem mabmbcmcdm, the formula that is induced is the following: $[s1:m; s2:a], [s1, s2, (N(s2)), s2]$. The first part of the formula is initialization. There are two subsequences, denoted $s1$ and $s2$. $s1$ starts with m , and $s2$ starts with a . The second part of the formula gives instructions for producing the sequence. The instructions are interpreted as follows: $s1$: write the current symbol of $s1$; $s2$: write the current symbol of $s2$; $(N(s2))$: change the symbol in $s2$ to the successor (N for next) of the current symbol; finally, $s2$: write the (new) current symbol of $s2$. The entire sequence is generated by repeating this routine as many times as necessary.

The problem solver constructs a formula as a hypothesis, based on the first letters of the given sequence, and tests the hypothesis with more of the letters. There are many different ways to form a sequence of letters, so in principle, the number of possible formulas is extremely large. To make the task manageable, some constraints have to be imposed. In Simon and Kotovsky's (1963) model, constraints are imposed on the generation of hypotheses. As in the focussing strategies that Bruner et al. (1956) identified for concept induction, hypotheses about the structure of a pattern are based on features of the stimulus, rather than being generated a priori. Furthermore, only a limited set of the possible hypotheses are ever generated, because the model only considers a small set of relations between elements and it is assumed that the sequence fits a specific form.

The model knows the alphabet of letters, both forward and backward. The relations that are recognized are identity and successor, I and N . The problem solver assumes that the sequence is periodic, an important structural characteristic.

The model begins by determining the period of the sequence. Periodicity can be discovered either by noting that a relation is repeated every n th symbol, or noting that a relation is interrupted at every n th position. In the problem mabmbcmcdm the periodicity is identified by noting that the relation I occurs between every third symbol, the m 's. Then the problem solver produces a description of the symbols that occur within the periods and relations between corresponding symbols in successive periods. For mabmbcmcdm the description requires two subsequences, one of which is just repetition of m ; the other starts with a and is incremented to produce the final term within the set of three symbols. The result of the process is a formula for producing the sequence, such as the one described earlier for the example problem.

Because the product of the inductive process is an explicit formula, sequence extrapolation can be considered as a problem of design as well as a problem of induction. Viewed in this way, the problem solver has available a set of symbols -- $s1, s2, s3$, (perhaps more), N , and the letters of the alphabet -- and has the task of constructing from these symbols. The feature of sequence extrapolation that makes it an inductive task is the criterion that the construction must satisfy. The criterion is that the formula should produce the sequence of letters that is given in the problem. In ordinary problems of design, such as anagram or cryptarithmic, the criterion is a general property rather than agreement with an arrangement of stimuli.

Simon and Kotovsky (1963) reported data on difficulty of solving 15 different sequence-extrapolation problems by two groups of human subjects and found fair agreement between the relative difficulty of problems for human solvers and for their program. A more thorough empirical study was conducted by Kotovsky and Simon (1973), who collected thinking-aloud protocols on problems with sequences under presented under panels that subjects lifted to see individual letters. The data were consistent with the model in important respects. Subjects, like the model, determined the periodicity of sequences and looked for relations between successive elements or between elements separated by a regular period. Representations of sequences induced by the subjects agreed with those that the model induced in a majority of cases.

There also were discrepancies, some of which involved relatively minor details of programming, but two of which revealed significant processes in humans not represented in the model. First, there was a closer integration in the subjects' performance than in the program's between discovery of the period of the sequence and induction of the pattern description. These are distinct phases in the model, whereas the human problem solvers used information in forming the pattern description that they had picked up during the phase of finding the period. Another discrepancy between human data and Simon and Kotovsky's simplest model was that in some problems, human solvers induced patterns with hierarchical structure, involving a single low-level description and a higher-level switch that transited between versions of the low-level structure. Hierarchical relation between levels of pattern description is a basic structural feature of sequential patterns that can play a dominant role in the induction process, as shown by Restle (1970).

IV.B.2. Grammatical Rules. Next, we consider induction of the grammar of a language. We discuss aspects of language acquisition that relate directly to general issues in the theory of problem solving. Thus, our discussion is selective, and does not fully represent the rich literature on processes of language acquisition, which deals with a much broader range of issues than we consider here.

In acquiring the grammar of a language, the materials presented to a learner include sentences of the language. The task is to infer a set of rules that can be used to parse sentences that are heard and produce sentences that are grammatical in the language. Thus, problem solving involves search in a space of possible syntactic rules. The space of stimuli includes the grammatical sentences that the learner hears, and the task is to induce the rules that characterize the structure of those sentences.

Human knowledge of the rules of grammar is implicit, in contrast to the explicit formulas that are induced in the sequence extrapolation task. This is seen in the facts that very young children have significant knowledge of grammar (e.g., Brown, 1973), and that adults know grammatical rules explicitly only if they have had special training. Because of the implicit nature of grammatical knowledge, the product of language learning is characterized as a set of procedures, rather than explicit formulas or other descriptions of structure. The procedures acquired by learners of a language enable them to produce and understand sentences that agree with the grammar of the language and to distinguish between grammatical and

ungrammatical sequences of words. We refer to such a set of procedures as knowledge of the grammatical rules, because the rules are built into the procedures. As with much procedural knowledge, an individual's knowledge of the rules in the form of procedures does not imply that he or she can state what the rules are.

There is evidence from both empirical studies (e.g., Moeser & Bregman, 1972) and theoretical analyses (e.g., Wexler & Cullicover, 1980) that grammatical rules are learned more easily if reference is provided for terms in the language. This indicates that in the space of stimuli for inducing a grammar, each sentence is paired with a situation that the sentence describes. The functions of situations in facilitating induction of grammatical rules probably include assisting in determining which words belong together in constituent units (cf. Morgan & Newport, 1981).

We discuss an analysis of language acquisition by Anderson (1975, 1977), as an example that describes a definite information-processing mechanism for acquiring knowledge of grammatical rules in the form of procedures. Anderson's system includes learning processes that show how semantic reference can facilitate the acquisition of grammar. Anderson's learning system, called LAS for Language Acquisition System, induces rules of grammar when it is given sentences in a language accompanied by the semantic objects that the sentences are about. For example, if the sentence, "The red square is above the small circle" is presented to LAS, there also is a semantic network that represents an object with the properties red and square, another object with the properties small and circle, and the relation above between the two objects.

LAS has a procedure, used in its learning of grammar, that identifies the objects in the semantic network that correspond to words in the sentence and forms a structure showing the relations among those concepts. This structure is used to determine constituent units of the sentence. In the example sentence, red and square are bracketed together, because they are properties of the same object, as are small and circle. The relational term above is at a higher level in the bracketing that LAS forms. The procedures that LAS acquires include rules for parsing noun phrases such as the red square and the small circle, and sentences of the form NP Relation NP. LAS also has a mechanism for generalization, so that similar structures eventually come to be parsed by LAS with a single rule, and some of these generalizations produce recursive parsing rules. The generalization process sometimes produces incorrect rules that are too general, and LAS also includes a mechanism of discrimination that restricts the application of its language-processing procedures.

Viewed as a problem-solving system, LAS conducts search in a space of procedures for producing and understanding sentences. (Note that we can also view LAS as designing or constructing these procedures.) LAS's use of the structure of situations provides significant constraints that are needed for the search. As in Simon and Kotovsky's (1963) model of sequence extrapolation, the constraints are applied to the generation of hypotheses. Processes for modifying the induced procedures are available; the system can generalize its procedures, which makes its performance more efficient, and it can add restrictions to the application of procedures when it is informed that use of a procedure has produced an error.

IV.C. Nonsequential Patterns

Now we discuss induction of patterns that are not sequential in character. We begin with a simple case. We discuss analogy problems in which one or two pairs of items are presented, related in some way. The task is to form another pair with the same relation. There have been extensive empirical and theoretical analyses of processes of solving analogy problems. We then discuss more complicated cases, involving induction of concepts in mathematics and of quantitative regularities and structures in scientific domains, for which the available analyses are primarily theoretical.

IV.C.1. Analogy Problems. The form of an analogy problem is A:B::C:D, where D is often a set of alternative items that can complete the analogy, with the subject required to choose one from the set. A and B are related in some way, and the correct choice is a D item with the same relation to C as B has to A. Solution of an analogy problem involves search in a space of relations for one that can be applied to both the A:B and the C:D pairs, or to one of the C:Di alternatives more successfully than any of the others. Analogy problems are used commonly in tests of intellectual ability. In factor-analytic studies, analogy problems contribute most to the factor of induction, the single best predictor of academic achievement (Snow, 1980).

Solution of analogy problems requires (1) a process for recognizing or analyzing relations between pairs of stimuli, that is, between the A and B stimuli and between C and each of the Di alternatives; and (2) a process that compares relations found for the A:B pair with relations found for the various C:Di alternatives and chooses the C:Di relation that matches best with an A:B relation. In the simplest case, the relation for A and B that comes to mind first also applies to one and only one of the C:Di pairs. When this does not occur, because relations :B found for A:B apply either to more than one C:Di pair or to none of them, some further analysis of the A:B pair is required. In such cases, A:B relations can be suggested by relations that are found in considering the C:Di pairs.

We discuss two processes for solving analogy problems. In one process, relations between pairs of items are based on information stored in the problem-solver's memory. Memory-based analogy problems include most verbal analogies, where solutions use relations between words that are stored in memory or are inferred from word meanings. In the other process, relations are determined by analysis of features of stimuli. When analogy problems are composed of geometric diagrams, relations between pairs of terms are found by comparing pairs of diagrams and identifying differences between the members of each pair.

Relations Based on Semantic Memory. Solutions to many verbal analogies are obtained by finding a relation between the A and B words based on their meanings stored in semantic memory, and then finding a similar relation between C and one of the Di pairs. Reitman (1965) formulated a model for verbal analogies based on activation of concepts in a semantic network. Reitman's model, called Argus, solves problems such as bear : pig :: chair : (foot, table, coffee, strawberry). Argus has knowledge of words in a network of relational connections; for example bear and pig are both connected to animal through the relation

superordinate. Activation and inhibition are transmitted through connections between units.

Argus can perform according to different strategies. In one strategy, the A and B terms are activated, and relations that become made active are noted; then C is active, and the Di alternatives are activated in turn. A goal is set for relations that are the same as the ones activated by the A:B pair. When a C:Di pair activates those relations, that Di alternative is chosen. In the example problem, after bear and pig are activated, their superordinate relations to animal become active, because these lie on a path between the activated terms. Then chair is activated along with the Di alternatives in turn, with the goal of finding active superordinate relations. This goal is achieved when table is activated, because both chair and table are connected by superordinate relations to furniture.

Strategic factors in analogy problems were demonstrated in an experiment by Grudin (1980). Grudin presented two kinds of analogy items: one, called standard items, where a salient relationship between A and B can be matched with one of the C:Di pairs; and the other, called nonstandard items, where there is no salient relation between A and B, but a relation between A and C matches one between B and a Di alternative. An example is the item bird:air::fish:(breath, water, swim) in standard form, which in its nonstandard version is bird:fish::air:(breathe, water, swim). The nonstandard problems are more difficult, as measured by the time required for a solution. However, if subjects can adapt their strategies to look for relations between A:C and B:Di pairs, the difficulty of nonstandard problems might be reduced. Grudin's sequence of problems included five-item sets that were either all standard or all nonstandard, followed by either a standard or a nonstandard problem. During solution of a set of nonstandard items, a shift in strategy could occur, involving more attention to the A:C and B:Di pairs. This would produce shorter times for nonstandard problems following nonstandard sets than for nonstandard problems following standard sets, and this result was obtained.

Thinking-aloud protocols in solution of verbal analogies were obtained in a study by Heller (1979; also described by Pellegrino & Glaser, 1982). Heller first presented the three terms of an analogy stem and asked the subject to think aloud, including a statement of any A:B relations and expectations about the answer that came to mind. Then four alternative answers were presented individually with the subject asked to judge whether each alternative was an acceptable answer, and why, and finally the complete problem was presented for a final choice.

Heller's findings were consistent with the general features of Reitman's (1965) hypotheses of solution strategies and of finding relations by activation of a semantic network. Strategic factors provide an interpretation of individual differences in performance, and the activation hypothesis is supported by a finding of variability in solution sequences.

Heller's major finding was a striking difference between groups of subjects in their adherence to the task constraints of analogy problems. The main constraint of an analogy is that the relationships between A:B and C:Di should correspond. If a subject chooses a Di response on the basis of a relation to C without regard to the correspondence of that relation to the A:B relation, then the analogy constraint has not been applied.

Subjects who had good overall performance mentioned the similarity or difference between an A:B relation and at least one of the C:Di relations on nearly all problems. In contrast, subjects with poorer overall performance were inconsistent in applying the constraint of a matching A:B and C:Di relation, frequently accepting answers based on a relation between Di and C, or with other terms in the analogy of a quite diffuse kind. To account for the differences among subjects in adherence to the task constraints, Heller proposed that individuals differ in the strengths of goals that are related to general solution strategies. In Reitman's model, this would correspond to the better subjects' having strategic goals that were activated more strongly, or to differences in the degree to which goals became inactive or were interfered with by other processes.

Heller's protocols also revealed considerable variability in the sequence of steps in solving the problems. In a majority of cases, subjects identified an A:B relation and then thought about C:Di alternatives in the context of that relation. There also were cases in which a relation between A and B came to mind as a subject thought about one or more of the C:Di relations. Such solution sequences occurred in about 20% of the problems for which subjects adhered to the analogical constraints. Reitman's assumption that relations are found by activation of a semantic network provides an interpretation of the variability of solution sequences, since activation of a relation in the context of a C:Di pair would facilitate its recognition for A:B in some cases where A:B did not elicit it.

Further information relevant to individual differences was obtained in a study by Pellegrino and Glaser (1982). Analogy items with single D alternatives were presented and subjects judged the items as true or false. Pellegrino and Glaser used an experimental and statistical method introduced by Sternberg (1977), in which the four terms are presented in sequence, with the subject making a response to request the addition of each term. The latencies of these responses are used to estimate the time for various component of the solution process, according to a general model. Each latency includes time to encode the new item. When B is presented, the latency includes time to infer one or more relations between A and B. When C is presented, the latency includes time to map A:B relations onto the C term. When D is presented, C:D relations are inferred and compared with the A:B relations. It was assumed that the comparison process could have three outcomes. The relations could correspond well, leading to a "true" response. The lack of correspondence could be so great that the subject would immediately reject the analogy and give a "false" response. The subject could judge that the correspondence was indeterminate and engage in a more extended analysis, possibly including review of the A and B terms to find new relations.

Four sets of items were used in the study. There were positive items, which were judged to be appropriate analogies, and negative items, which were judged to be inappropriate. Within each of these sets, there were items in which the C and D terms were strong associates and other items in which the C and D terms were not associates. A weak C-D association for a positive item, or a strong C-D association for a negative item, was expected to make the item more ambiguous and increase the frequency of extended analyses in the final component of the solution process. The results supported this expectation; estimates of the proportion of

problems with extended analyses were higher for weakly associated positive items (.55) than for strongly associated positive items (.23), and for strongly associated negative items (.19) than for weakly associated negative items (.07). A similar correlation of item difficulty with time spent in the final stage of solution was obtained by Barnes and Whitely (1981).

Pellegrino and Glaser's major finding was that the frequencies of an extended analysis were correlated with the subjects' overall ability in the analogies task. The subjects were college students divided into groups with relatively high and relatively low scores on a standard analogies test. The estimates of time for the various information-processing components were generally longer for the low-ability subjects. But the most striking difference was in the frequency of engaging in an extended analysis, which was more than twice as high for the low-ability than for the high-ability subjects. Pellegrino and Glaser concluded that the low-ability subjects more frequently arrived at the final stage of processing an analogy with an inadequate representation of the relations among the other terms, and therefore had to reconsider the A, B, and C terms more frequently. (A similar difference in the solution process was found by Snow, 1980, in spatial reasoning tasks in which the items are diagrams and reexaminations of terms can be observed by recording eye movements.) In verbal analogies, this difference in processing could be due to differences in the information in semantic memory, differences in the activation process, or differences in strategy with low-ability subjects more likely to want to see the final term to facilitate recognition of A:B relations. This conclusion is consistent with Heller's finding that students with lower ability in analogies often choose responses that violate the constraint of an analogy problem. They frequently lack a response that satisfies the constraints, and are likely to choose a response on some other basis.

In Reitman's (1965) model of verbal analogy solution, relations are relatively discrete components of semantic memory. This characterization probably is correct for most verbal analogies, but there are cases in which it does not apply. An example was studied by Rumelhart and Abrahamson (1973), who studied solution of verbal analogy problems in a single semantic domain, the names of animals.

Analogies composed of animal names have two properties that are different from most verbal analogies. First, they depend on more than one relation, and the relations are combined somehow in solving the problem. Second, the relations differ in degree, rather than just being present or absent.

An example that illustrates multiple relations is the following: rabbit:sheep::beaver:(tiger,donkey). Donkey seems the better answer, perhaps because while a relationship involving size is similar for beaver:tiger and beaver:donkey, and both are similar to the size relation for rabbit:sheep, there also is an additional difference for beaver:tiger -- tigers are ferocious while beavers are not, and thus the beaver:donkey pair matches the rabbit:sheep pair better, which also lacks a difference in ferocity. The graded nature of relations is illustrated by rabbit:beaver::sheep:(donkey,elephant). Donkey seems the better answer. The judgment seems to depend mainly on the sizes of the animals, and

beavers are larger than rabbits, but the difference seems not to be large enough to make sheep:elephant seem appropriate.

To represent differences of graded magnitudes that can be combined easily, it is convenient to use a spatial representation. In such a representation, the dimensions of the space correspond to salient ways in which items differ from each other. Each item is located at a point in the space. The coordinates of the point correspond to the values that the item has on each of the dimensions.

A spatial representation of a set of items can be obtained by presenting pairs of the items to subjects and asking them to judge how similar the members of each pair are to each other. These judgments of similarity are used as estimates of the distances between pairs of items, and items are located in the space so that the distances between points are as close as possible to the estimates obtained in the experiment. In the method of choosing the spatial representation, called multidimensional scaling, an attempt is made to represent the items in one dimension; if that is unsuccessful two dimensions are used, and so on until a space is found with the points located so that interpoint distances agree satisfactorily with the similarity judgments given by subjects.

Henley (1969) obtained judgments of similarity for pairs of animal names, and obtained a spatial representation with three dimensions: size, ferocity, and a third dimension that probably involves a mixture of attributes, including similarity to humans. These results were used by Rumelhart and Abrahamson (1973) in their study of analogy problem solving. The relation between two items A and B corresponds to the vector that connects the points for A and B in the spatial representation. The vector represents the combination of differences in the three dimensions between the two items; for example, the vector from beaver to tiger represents a moderate increase in ferocity, a large increase in size, and very little difference in "humanness." In Rumelhart and Abrahamson's model, to solve an analogy, $A:B::C:(D1,D2,D3,D4)$, the $A \rightarrow B$ vector is translated to C, and the probability of choosing each of the D_i alternatives is a function of its distance from the ideal point defined by the end of the vector. In one experiment, the model provided accurate predictions of the frequencies of subjects' rankings of response alternatives in analogy problems. In another experiment, fictitious animal names were locations in the spatial representation. These fictitious names were used in analogy problems for which subjects received feedback, and the subjects induced features of the fictitious animals, responding appropriately to new analogies involving their names.

Figure 12 here

Relations Based on Feature Analysis. In a geometric analogy problem, the terms are diagrams that differ in various ways. An example is in Figure 12. The best answer apparently is D2. A and B are related by deletion of the dot and moving the rectangle from inside the triangle to the left of the triangle. C and D2 are related in a similar way: the dot in C also is deleted, and the Z is moved from inside the segment of the

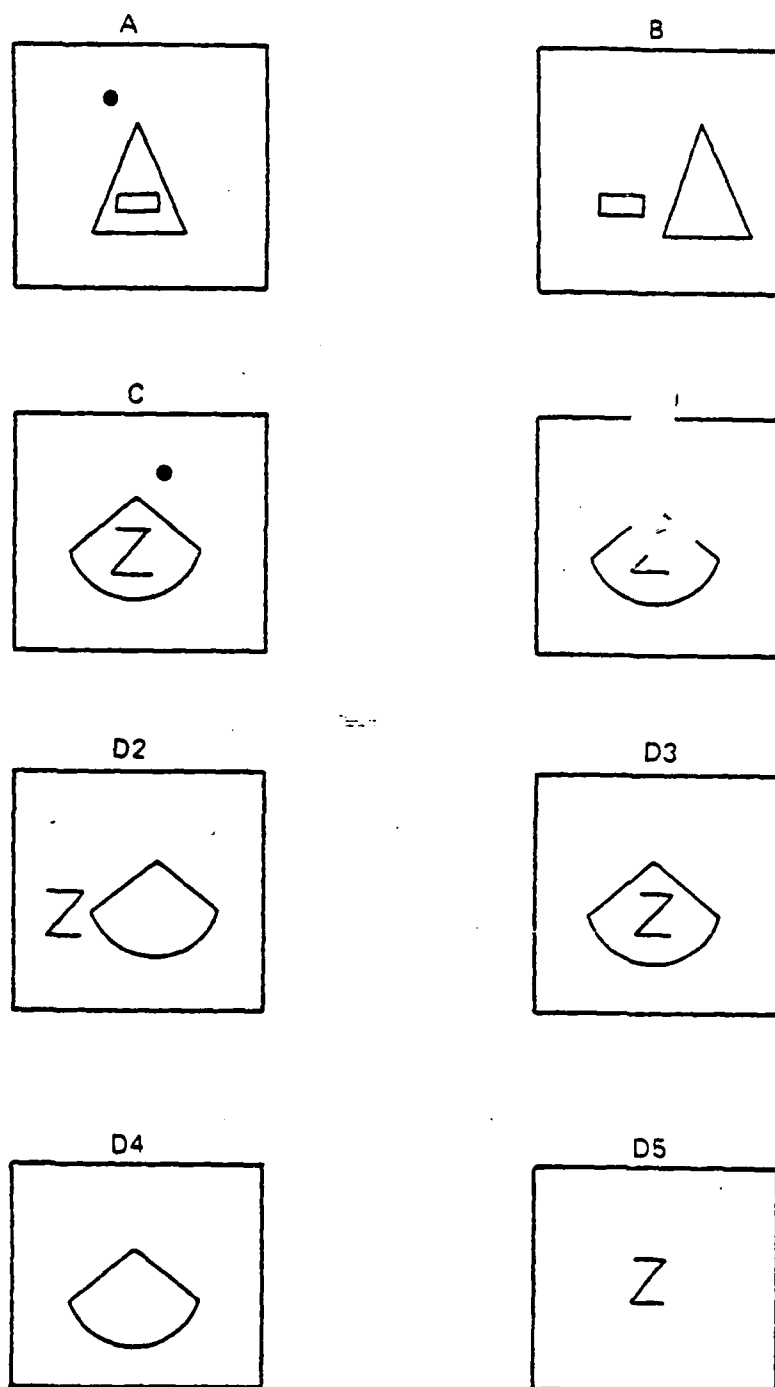


Figure 12. A geometric analogy problem (from Evans, 1968).

circle to the left of the segment.

As Figure 12 illustrates, the relation between a pair of diagrams can have several parts, corresponding to components of the diagrams that differ. Some of the differences can be quantitative, for example, the amount of rotation of a component or the amount by which the size of a component is increased or decreased. In the domain of analogy problems involving animal names, these characteristics of composite and quantitative relations make a spatial representation of items a reasonable one. A spatial representation is not economical for geometric analogies, because there are too many ways in which diagrams can differ. In the domain of animal names, a satisfactory approximation is that all pairwise relations can be characterized by differences on three dimensions, but the domain of geometric diagrams does not have such a simple structure.

In geometric analogies, relations are found by examining features of the diagrams, rather than by retrieving information from memory, as with verbal analogies. A model of solving geometric analogy problems, therefore, has two components: one component that analyzes diagrams and identifies relations between them, and another component that compares the relation of A:B with relations of the C:Di alternatives and chooses the best match.

Evans (1968) developed a model that solves geometric analogy problems. The program is given descriptions of the diagrams that specify the locations of straight lines, curved lines, and closed figures. From these descriptions, relations among components are derived; for example, that one figural component is inside another, or is above it in the diagram.

The model then compares its representations of the diagrams in pairs and forms descriptions of relations between the members of the pairs. These relations are in the form of transformations -- that is, changes in one diagram that would make it the same as the other diagram in the pair. For example, a component in one diagram might be removed, or a component might be added or a component might be changed in size or rotated, or the relative positions of two components might be changed, say, by moving one from inside the other to above the other.

The relation between A and B is then compared with the relations between C and each of the Di alternatives. This comparison involves matching components of A with components of C and determining which of the transformations in the A:B relation also occur in the C:Di transformation. The Di alternative is chosen for which the greatest number of transformations can be made to correspond.

Evans (1968) developed his model as a project in artificial intelligence, rather than as a simulation of human problem solving, but even so, the model has features that seem plausible as psychological hypotheses. One such feature is a suggestion that problems with more complex diagrams or relations between diagrams should be more difficult for humans to solve. In the model, diagrams are more complex if they have more components, and relations are more complex if there are more transformations, that is, if there are more changes in components between diagrams that are related. These factors were varied in an experiment by Mulholland, Pellegrino and Glaser (1980), and both had significant effects.

Problems that had diagrams with more components and problems with more transformations required longer times for solution.

In human solution of geometric analogy problems, we should expect some of the same characteristics of performance that have been observed in solution of other analogy problems. An important factor in verbal analogy problems, discussed above, is the processing required when the subject's representation of the A:B relation and the C:D_i relations are not sufficient to provide a determinate answer and further processing is needed. Findings by Sternberg (1977) show that this factor is important in geometric analogy problems as well. Sternberg measured the time to solve problems presented after part of the problem had been shown, enabling part of the processing to occur. He used the differences between conditions as estimates of the times for components of the solution process. In comparing subjects with differing levels of general reasoning ability, Sternberg found a large difference in the time required to process the C:D_i alternatives in geometric analogy problems, with much of the difference attributed to a process of comparing alternatives when prior processing has not provided a unique solution.

IV.C.2. Inductive Problems in Mathematics and Science. Cognitive analyses have been developed in the form of computer programs that invent new mathematical concepts, based on properties of examples, and that induce formulas and structures from data in scientific domains. We briefly discuss three models: one that invents new mathematical concepts, one that induces formulas from sets of quantitative data, and one that induces molecular structure from data of mass spectroscopy.

Invention of Concepts in Mathematics. A program called AM (Lenat, 1982) generates examples of concepts that it knows and develops new concepts, based on properties of the examples. The main domain in which AM was run is elementary mathematics. AM was given initial concepts involving sets and developed a variety of concepts involving numbers. For example, AM developed concepts of addition and multiplication, developed the concept of primes, and arrived at a conjecture that every number is the product of a unique combination of prime numbers.

It is useful to compare AM's task to the standard experimental task of concept induction, such as that studied by Bruner et al. (1956). In standard concept induction, a set of examples is provided by the experimenter, with some positive examples and some negative examples determined by a rule, and the subject's task is to induce the rule. Hypotheses are generated by the subject and tested with information about further examples until the correct concept has been found. Each hypothesis that is generated is itself a concept, in the sense that it provides a rule for classifying the stimuli. The main problem-solving work is in finding which rule is correct.

AM's task is not defined as well, in two respects. First, the examples are not provided by an experimenter, but rather are produced by AM. Second, AM does not have a specified criterion of correctness for the concepts that it generates. Instead, AM evaluates its concepts by some criteria of importance, based in part on how easy it is to generate examples of the concept.

AM's knowledge of concepts is organized with a set of facets, including some that are standard for semantic networks, such as generalizations, specializations, and examples, and others that are especially useful in the domain of mathematics, such as objects that are in the domain or range of a function. Facets also hold procedural information, such as methods for testing whether an object is an example of the concept. AM's reasoning activity is organized as a set of tasks, each involving a concept and one of its facets. Examples of tasks include filling in examples of a concept or forming a generalization or a canonical representation of a concept. Tasks that are proposed are placed on an agenda, and choice of a task to perform is based on an evaluation of the reasons for the task, including the importance of concepts for which the task would contribute new information. Heuristics that contribute to the development of new concepts include efforts to form a more general concept if an existing concept has very few examples, and to form new representations that clarify the relations between concepts.

We note that AM does not really do mathematics in the usual sense. It has no concept of deductive consequence, and thus does not develop a body of concepts and principles with a formal structure. Even so, it provides an example of a system that goes well beyond the knowledge that it is given initially, moving into a conceptual domain that is quite distinct from that of its initial concepts.

Table 10 here

Inducing Quantitative Regularities. A system called Bacon induces formulas from numerical data (Langley, 1981; Langley, Bradshaw & Simon, 1980). The data are values of some variables that are controlled and other variables that are measured; a simple example is in Table 10. The goal is to find a formula that describes the relation between the variables, in this case distance and time. The two components of the problem space are the subspace of stimuli, the set of data, and the space of structures, the set of formulas with the variables that are included in the data.

A simpler approach than Bacon's is adequate for relatively simple induction problems. This simpler approach tries to fit alternative formulas that are known in advance. For example, for Table 10, a linear function can be tried, and the discrepancy that is noted shows that there is positive acceleration. This suggests trying a quadratic formula, which fits the data. Generate-and-test methods of this kind have been analyzed by Huesmann and Cheng (1973) and by Gerwin (1974), with supporting experimental data.

The task of inducing formulas can become unmanageable for a simple generate-and-test method if there are several variables that can be related in complex ways. For example, Bacon is able to induce Coulomb's Law, relating electrical force to the charges on two bodies and the distance between them: $f = q_1 q_2 / d^2$; and a formula for the electrical current in a wire connected to a battery and a metal rod, depending on the temperature differential of the bar, the internal resistance of the battery, and the length and diameter of the wire: $I = T / (R + L/D^2)$. The set of formulas that includes these is extremely large, and it seems unlikely that simple equation fitting would be an effective method for inducing formulas of this

Table 10

Data for a Simple Induction Problem

Time	Distance
1	0.98
2	3.92
3	8.82
4	15.68
5	24.50

complexity.

Bacon's search method uses properties of the data to guide formation of hypotheses. We have discussed other induction systems with this property, including the concept-induction strategy of focussing, described by Bruner et al. (1956), the method of inducing patterns in letter sequences studied by Simon and Kotovsky (1963), and AM's heuristics for generating new concepts based on properties of examples. Bacon's heuristics involve properties of quantitative data and thus differ, as one would expect them to, from the heuristics of other systems such as AM, where the data involve categories of examples and sets of defining features. Bacon's use of data has the further interesting feature of creating new data in the process of evaluating its hypotheses. In evaluating a hypothesis, Bacon calculates values of a new function of available data, and if the hypothesis does not succeed, those values become part of the data available to Bacon for further problem solving. Thus, an attempt to solve the problem may be unsuccessful, but it leaves new results that may be instrumental in a later attempt that succeeds.

Bacon's basic method is to search for a function of data that gives constant values across experimental conditions. As an example, the formula for the data in Table 10 is $d = kt^2$, where k is a constant; the form in which Bacon discovers the law is $d/t^2 = k$.

Bacon uses heuristic rules to form hypotheses, consisting of functions of variables in its data base, that might give constant values. For example, if two quantities increase or decrease together, Bacon forms their ratio as a new quantity to be considered. If one variable decreases as another increases, Bacon forms their product as a new quantity. These heuristics, and another that forms linear functions of variables, enable Bacon to induce relatively complex functions. (The first two are sufficient for the problem in Table 10. First, note that t and d increase together, and form the ratio t/d . This decreases with t , so form the product t^2/d . This quantity is constant across the observations.)

Some other heuristic methods are also used, including definition of "intrinsic variables" as properties of objects that are associated with constant values of quantities, and attempts to find a common divisor for values of intrinsic variables that have been induced. These heuristics enable induction of properties such as the resistances of different wires from measurements of current, and the atomic and molecular weights of chemical elements from data about weights and volumes of elements and compounds involved in chemical reactions.

As mentioned previously that induction problems can also be understood as problems of design, especially when the structures that are induced are expressed explicitly as formulas. This view is particularly appropriate in the case of Bacon's induction of formulas. Consider the task as construction of a formula using symbols for the variables in the problem. Bacon's heuristics then are rules for forming combinations of the symbols that may satisfy the problem criterion. If a formula does not solve the problem, it may provide part of the formula that is needed. Thus, the process of search with construction of partial solutions, characteristic of design problems, provides an appropriate characterization of Bacon's process of induction.

Bacon is not intended as a complete simulation of cognitive processes in scientific research, where hypotheses about causal mechanisms often play a critical role in decisions to measure variables or to examine quantitative relationships. Even so, it provides a demonstration that quite simple heuristics are sufficient to produce quite complex inductive conclusions from quantitative data, and it is reasonable to suppose that these heuristics correspond to significant components of complex scientific reasoning.

Inducing Molecular Structure. Another scientific task that has been investigated is induction of the molecular structure of organic compounds. A system called Dendral induces molecular structure from data in the form of mass spectra (Lindsay, Buchanan & Feigenbaum, 1980). A mass spectrum is a set of quantities of the fragments of various sizes that are produced when molecules of a substance are bombarded by electrons.

Like AM and Bacon, Dendral performs induction using heuristic search. An important difference is that Dendral uses search heuristics that are based on principles that are specific to the domain of organic chemistry, whereas AM's methods apply in any domain with a structure of categorical concepts, and Bacon's methods can be applied to any quantitative data.

Dendral's method of induction has three main stages. First, the chemical formula of the compound is inferred from features of the mass spectrum. Then hypotheses about molecular structures are generated with constraints based on knowledge of the class of compounds that the substance belongs to. Finally, the hypotheses are tested by comparing their implications with the quantitative details of the mass spectrum, and the hypothesis is chosen that provides the best agreement with the data.

The data used to infer the chemical formula are the peaks in the mass spectrum. The largest mass represented probably is the mass of the molecular ion, or may be smaller than the molecular ion by one fragment. Differences between peaks usually correspond to the masses of fragments that are broken off in the bombardment. Dendral uses the value of the largest peak and the interpeak distances, along with knowledge of chemistry, to infer one or more chemical formulas that are consistent with the spectrum.

Dendral's next task is to generate possible molecular structures, with the ions in the formula arranged in various ways that are consistent with known possible arrangements. There are many millions of possibilities for most problems, so Dendral formulates constraints based on knowledge of the class of compounds that the sample belongs to. With the constraints, Dendral constructs hypotheses about molecular structure with a method that first determines the maximum number of rings in the structure, then constructs the possible partitions of ions into rings and remaining components, and finally constructs the possible structures for each possible partition.

Finally, Dendral tests its many hypotheses, using the quantitative details of the mass spectra. In the different hypothesized structures, different components are separated by different numbers of bonds; therefore, there are differences in the likelihoods that they will occur together in a fragment. Assuming that fragments are produced by breaking

one or two bonds at once, predictions are made about the relative amounts of material to be found at each peak in the spectrum, and the structure that fits the data best is chosen.

Note that Dendral's task, like Bacon's, involves constructing an explicit formula to represent the structure that it induces. Thus its method can also be considered as solving a problem of design, where the materials for the construction are symbols that represent the atomic components of chemical compounds, and the chemical knowledge that it uses constrains the search for an arrangement of those materials that satisfies the criterion of agreement with the mass spectrum.

IV.D. Diagnostic Problem Solving

We conclude this section by discussing problem solving that involves troubleshooting in electronics and diagnosis in medicine. In these tasks, the problem solver has a space of stimuli consisting of one or more symptoms and further information that can be obtained by performing tests. The space of structures is a set of possible causes of the symptoms, faulty components in electrical circuits or disease states in the case of medicine.

In addition to its characteristics of inductive problem solving, diagnostic problem solving also has components of operational thinking, because it is based on the goal of curing the patient's illness or repairing the device. Thus, the information and conclusions in the diagnosis are directed toward making a decision about a remedial treatment that should be applied.

IV.D.1. Troubleshooting. The task in troubleshooting is to determine which of the many components of an electronic system is operating incorrectly, causing the system to function improperly. There may be more than one fault, but it simplifies the problem greatly to assume that there is a single fault in the system.

In a general way, troubleshooting resembles the task of inducing categorical concepts when the subject chooses the stimuli for which information is given. In concept induction, the problem solver obtains information by asking whether a specific stimulus is a positive or a negative instance. In troubleshooting, information is obtained by taking readings of voltage or current at specific locations in the circuit. In both cases, there are many possible hypotheses to be considered, but the set of possibilities can be specified — in concept induction, it is the set of logical combinations of the stimulus attributes, and in troubleshooting, it is the set of possible faults of components. These similarities of the tasks are correlated with an important resemblance in the effective methods of working on the problems. The focussing strategy in concept induction uses information obtained about instances to eliminate classes of hypotheses, rather than considering each hypothesis individually as is done in the less effective scanning strategy (Bruner et al., 1956). Similarly, in troubleshooting, an important component of strategy is to conduct tests that will enable elimination of sets of possible faults from consideration. Use of this strategy is enabled by general knowledge about electronic components as well as by knowledge of the specific circuit in the problem, as we discuss below. This requirement of knowledge to support

the process of induction is analogous to the role played in concept induction by knowledge of the alternative logical forms (conjunction, disjunction, etc.) and the truth-table combinations that correspond to them (Dodd et al., 1971), although the knowledge required in troubleshooting is considerably more elaborate.

A model of troubleshooting is included in a system called Sophie that provides computer-based instruction for trainees in electronics maintenance (Brown, Burton & deKleer, 1983). The troubleshooting system provides a model for the student to observe in learning how to diagnose faults in a circuit. The student can specify a fault in the circuit, and Sophie then can solve the problem of diagnosing the fault, performing a series of tests to obtain readings of current or voltage at various points in the circuit, forming hypotheses about the fault, and eventually arriving at a decision about it. Sophie has general knowledge about electronics and an explicit representation of strategy that enables it to provide explanations to students for tests that it is performing, regarding both principles of electronics and the strategic purposes of its activity. Sophie's troubleshooting knowledge is also used to evaluate the problem-solving performance of students, by providing a series of problem-solving steps that can be compared with the steps taken by students.

Sophie's knowledge for troubleshooting has four main components: two components of electronics knowledge, a component of knowledge for making specific inferences and a component of strategic knowledge. One component is general knowledge about electronics in the form of "experts" that have information about characteristics of different kinds of electronic components such as resistors and diodes. These experts can use data obtained from readings to calculate values for other variables, assuming normal functioning of components of the circuit; these inferred values then can be compared with actual readings of those variables.

A second component of Sophie's knowledge is information about the specific circuit that is used for instruction. The circuit is represented hierarchically as a set of modules with submodules and components. Possible functional states of each module and component are represented, including normal functioning and possible fault states. Experimental evidence obtained by Egan and Schwartz (1979) is consistent with a hypothesis that human electronics experts represent circuits in ways similar to Sophie's. Egan and Schwartz showed that experts encode information from circuit diagrams rapidly, similar to performance by experts in other domains such as chess (see Section III.B), and that functional modules made up of components that are spatially contiguous in the diagram play an important role in the performance.

A third part of Sophie's knowledge involves specific actions that occur during troubleshooting. This knowledge is in the form of rules for making inferences about the states of modules and components of the circuit. Readings are used to eliminate hypotheses about faults by showing that a module is functioning normally, and for propagating inferences in the hierarchical representation; for example, if a component is faulted, then all the modules that contain that component must also be faulted.

The fourth component of knowledge is Sophie's strategy, a breadth-first search method with backtracking. Sophie considers all the possible states that can occur, according to its representation of the circuit, and eliminates possible fault states on the basis of readings that are consistent with normal functioning. It assumes normal functioning of components until there is a reading that conflicts with that assumption; however, it keeps a record of the assumptions used in its inferences, and if information contradicts an assumption made earlier, inferences based on that assumption are revised.

IV.D.2. Medical Diagnosis. In medical diagnosis, as in troubleshooting, a system -- in this case, a human body -- is functioning improperly, and the inductive task is to infer the cause of the malfunction. Also as in troubleshooting, the purpose of the diagnosis is to determine a treatment that can remedy the malfunction, and the diagnostic activity is conducted in a way that provides information relevant to choosing a treatment.

Several systems have been developed that solve diagnostic problems in various domains of medicine, including diagnosis of infectious agents and prescription of antibiotics (Shortliffe, 1976), prescription of digitalis therapy for cardiac patients (Silverman, 1975), and diagnosing and prescribing treatment for varieties of glaucoma (Weiss, Kulikowski & Safir, 1978). (For a review, see Ciesielski, Bennett & Cohen, 1977.) We discuss one system, Caduceus, which performs general diagnosis. We also discuss some empirical studies of diagnostic problem solving by physicians with varying amounts of training and experience.

A Model of Knowledge for General Diagnosis. Knowledge used in general medical diagnosis has been investigated in the context of a model named Caduceus (Miller, Pople & Myers, 1982; Pople, 1982). The knowledge that Caduceus has for diagnosing diseases is similar in important ways to the knowledge used by Sophie for diagnosing faults in electronic circuits. It is hierarchical in form, enabling systematic search in the space of hypotheses. Internist also has rules that infer hypotheses based on symptoms and test results and that propagate inferred information using the hierarchical structure of its knowledge.

Caduceus's knowledge about diseases is of two kinds, organized in two separate but related graph structures. One of these, called a nosological graph, provides a taxonomy of diseases based on the organs of the body that are involved and on etiological factors. This graph provides groupings of diseases based on their manifestations. The other knowledge structure, called a causal graph, contains information about disease states and processes. The causal graph contains technical concepts of pathology that refer to states of disease, such as cardiogenic shock.

Caduceus has the goal of identifying one or more disease entities that provide a complete explanation of a set of symptoms and findings in the case. Subproblems are formulated from findings that are not yet integrated in an explanatory network; these constitute diagnostic tasks that are generated by the system. Identification of the disease depends mainly on the nosological graph; this hierarchical structure is used in a top-down search to narrow the possible disease entities. The information about disease states and processes in the causal graph provides links between

hypothesized disease entities and the specific symptoms and test results that are available. Caduceus concludes its diagnostic analysis when an explanatory network has been developed that includes all the available symptoms and findings.

Empirical Studies of Diagnostic Performance. An extensive study of performance in diagnostic problems was conducted by Feltovich (1981; also described in Johnson, Duran, Hassebrock, Moller, Prietula, Feltovich, & Swanson, 1981). The results consistent were the general properties of the Caduceus model. They also provide information about characteristics of knowledge for diagnosis at different levels of experience and expertise. Feltovich obtained problem-solving protocols for cases in pediatric cardiology from individuals varying in experience from fourth-year medical students who had just completed a six-week course in pediatric cardiology to two professors who had more than 20 years experience in the subspecialty. Information from five cases was presented serially and the physicians gave their hypotheses and other thoughts about the cases, attempting to arrive at a correct diagnosis.

The performance of experts indicated knowledge that differed from that of novices in several ways, consistent with the general features of expert knowledge discussed in Section III.B. The major difference was that experts had more integrated knowledge about diseases. Experts also had more detailed knowledge of variations of disease states and more precise knowledge of relationships between diseases and symptoms. This was indicated in the performance of one advanced expert by his mentioning groups of hypotheses that were supported by the findings presented first, with later information used to narrow the range of possibilities. The other advanced expert used a more depth-first strategy, proposing a likely hypothesis based on preliminary findings, but modifying the hypothesis in a flexible way when later evidence provided counterindications. The knowledge of novices was primarily in the form of a few specific disease forms used in textbook cases. Novices responded to early evidence by proposing reasonable hypotheses, but were less likely to recognize the significance of later evidence and change their hypotheses when this was indicated. The sets of hypotheses mentioned by novices during problem solving were significantly smaller than those of the experts.

Similar conclusions regarding expert knowledge for diagnosis were supported in a study of expert and novice radiologists (Lesgold, Feltovich, Glaser & Wang, 1981). Lesgold et al. found that in reading x-ray films, experts generated representations in a three-dimensional system, using salient features to generate initial hypotheses that were refined or modified on the basis of more detailed features. Knowledge for recognizing features associated with abnormalities appeared to be well integrated with general knowledge of anatomy. The integration of experts' knowledge was evidenced by their ability to use features noted early as constraints on later interpretations (cf. Stefik, 1981). Interpretations of novices (in this case, first-year residents in radiology) depended more on finding an explanation for a few features, with a tendency for other details to be assimilated to the initial hypothesis rather than used to generate alternative hypotheses or modifications.

It is important to note the close similarity of conclusions from these studies of expert diagnosticians in medicine and the studies of expert performance in other problem-solving domains, especially physics and chess. Present findings indicate that a major source of expert performance is the expert's ability to represent problems successfully, and that this results from the expert's having a well integrated structure of knowledge in which patterns of features in the problem are associated with concepts at varying levels of generality, enabling efficient search for hypotheses about the salient features of the problem that cannot be observed directly as well as methods and operations to be used in solving the problem.

V. Evaluation of Deductive Arguments

The relation between human reasoning and formal logic has long been a subject of discussion and debate, and for some decades, a subject for experiment as well. It is generally agreed that human "logical" reasoning does not always conform to the laws of formal logic. Formal logic is a normative theory of how people ought to reason, rather than a description of how they do reason. It is important, then, to develop a descriptive theory of human reasoning to compare and contrast with the logic norms.

Experiments aimed at developing a theory of human reasoning have mostly set tasks of judging the correctness or incorrectness of formal syllogisms. These tasks require application of the rules of deductive argument which are special in some ways, and correct performance depends on the subject's knowledge and use of the technical rules of formal deductive inference. However, the processes used in these tasks do not differ in any fundamental way from those involved in problem solving in other domains. Psychological analyses provide no basis for a belief in deductive reasoning as a category of thinking processes that differ from other thinking processes, other than in the special set of operators that are permitted in rigorous deductive arguments. As Woodworth put the matter, "Induction and deduction... are distinguished as problems rather than processes" (Woodworth, 1938, p. 801).

We discuss studies of two tasks. First, we discuss propositional and categorical syllogisms, which present arguments in the sentential and predicate calculus. Subjects frequently make errors in evaluating these syllogisms, and research has attempted to explain why the reasoning process differs from correct logical inference. Second, we discuss linear syllogisms, which present arguments that depend on transitivity of order relations. Subjects make the transitive inferences in these tasks without difficulty, and psychological analyses have focused on the cognitive representation of information in the syllogisms.

V.A. Propositional and Categorical Syllogisms

Subjects in experiments on propositional or categorical syllogisms are asked to judge the validity of arguments such as the following (invalid) propositional syllogism:

If I push the left-hand button, the letter T appears.

I did not push the left-hand button.

Therefore, the letter T did not appear.

The major premise states what will happen if the button is pushed. It says nothing about what will or will not happen if the button is not pushed. Hence the conclusion does not follow from the premises. Yet in a typical experiment (Rips & Marcus, 1977) a fifth of the subjects accepted this as a valid syllogism.

Categorical syllogisms in the predicate calculus involve statements containing the terms some, all, and no. An example of a (valid) categorical syllogism is

Some jewels are diamonds.

All diamonds are valuable.

Therefore, some jewels are valuable.
Again, human subjects make frequent mistakes in judging whether certain kinds of categorical syllogisms are valid. For example, subjects are very likely to mistakenly judge that the following argument is a valid syllogism (Johnson-Laird & Steedman, 1978):

Some As are Bs.
Some Bs are Cs.
Therefore, some As are Cs.

In experiments on syllogistic reasoning, the type of syllogism presented is most commonly taken as the independent variable, and the numbers of subjects making errors on syllogisms of different kinds is measured. By comparing the error rates for different kinds of syllogisms, the experimenter seeks to formulate and test hypotheses about the cognitive processes that subjects use to make such syllogistic judgments.

For example, subjects will often accept, "No As are Bs and no Bs are Cs, therefore no As are Cs;" while they will almost always reject, "No As are Bs and no Bs are Cs, therefore all As are Cs." Yet both syllogisms are equally invalid. Such errors of reasoning have sometimes been attributed to an "atmosphere effect." In the example above, "no" appears in the premises, therefore "no" is more acceptable than "all" in the conclusion (Woodworth & Sells, 1935). Alternatively, some investigators have claimed that the reason for these errors is that the quantifiers and connectives, all, some, no, if...then, and, or, do not have the same meaning in natural language as they do in formal logic (Braine, 1978). According to this hypothesis, since the experimenter judges the correctness of answers by their conformity to the rules of formal logic while the subjects are using the natural language meanings, errors will be made when the two kinds of meaning diverge.

Errors and latencies in reasoning tasks depend not only on the form of the syllogism, but also on whether or not it has meaningful content (Wilkins, 1928). Thus, subjects may respond differently to the syllogism, "If some As are Bs and some Bs are Cs, then some As are Cs," and the syllogism "If some birds have blue eyes and some blue-eyed creatures are human, then some birds are human."

In general, subjects' error rates are lower when syllogisms have meaningful content, but there is an important class of exceptions. Subjects often reject valid syllogisms when the conclusions are contrary to facts known to them. "If all horses have four feet and all fish are horses, then all fish have four feet," may be rejected by subjects who know that fish are footless. The rate of rejection rises when subjects react emotionally to the conclusion. "If drug addiction is a disease and diseases should not be punished, then drug addiction should not be punished," will more likely be rejected by subjects who support strong measures against drug usage than those who do not (Janis & Frick, 1943; Lefford, 1946). Conversely, subjects often accept invalid syllogisms when the conclusions are consistent with their knowledge about the world or their preferences.

All of these findings must be stated as "tendencies," since many subjects who makes errors on some syllogisms of a certain form do not make such errors consistently. Moreover, there are large individual differences among subjects. For example, subjects who have had training in formal logic generally make fewer errors -- not surprisingly -- than subjects who have not had such training.

While human syllogistic reasoning conforms to some broad generalizations of the sorts that have been mentioned already, the findings derived from experiments are complex and confusing. In the past several years, a few investigators have sought to cut through the confusion by creating models of the inference process or some components of it. The attempt to create such models has revealed features of the reasoning task that had not been entirely obvious.

First, any one of a wide range of strategies might be used by subjects to solve the problems, and there is no reason to believe that all subjects use the same strategies. Subjects who reason by vague verbal analogies could succumb to the atmosphere effect, while other subjects who create semantic images of the propositions and reason by operating on those images might make quite different errors. (Certain syllogisms might require the creation of images more complex than a subject could handle in memory.) Subjects' knowledge of logical inference can be embedded in formal axioms or in inference rules, with different consequences for the likelihood of error. The axioms that define connectives or the inference rules might conform to some natural logic that deviates from the formal logic of the textbooks.

Several quite successful recent modelling efforts have used the idea that evaluation of syllogisms is a form of problem solving similar to that discussed in Section II.A. Using a set of inferential operators, the subject attempts to confirm the conclusion working from the premises, and accepts the conclusion if this problem-solving effort succeeds. The process typically used by subjects differs from the task of finding explicit proofs in that the inferential operators are not expressed overtly, and of course need not correspond completely to the rules of formal logic.

Models of evaluating propositional syllogisms have been formulated by Osherson (1975), Braine (1978), and Rips (1983). These models are based on the concept of natural deduction, discussed by Gentzen (1935/1969). A system of natural deduction is a form of production system. Rules for making inferences specify conditions in the form of patterns of propositions, and when a pattern is matched in premises, the inference is made. The models account for performance by postulating sets of inference rules assumed to be used implicitly by subjects. Rips also formulated a specific process of applying the rules and forming representations of the derivation. An interesting feature of Rips's formulation is the inclusion of suppositions that provide a backward-chaining component in the search process. A syllogism is judged valid if the system can generate a derivation of the conclusion from its inference rules.

The idea that sentential syllogisms are evaluated by natural deduction provides an interpretation of many of the kinds of errors that occur in syllogistic reasoning. Because it is an informal reasoning system, it is not surprising that it is susceptible to influence by general knowledge and affect. One would expect performance to be improved if subjects were taught a more explicit procedure for verifying the applicability of inference rules in evaluating syllogisms, and this result was obtained in the domain of geometry proofs in a study by Greeno and Magone (described in Greeno, 1983).

Models of reasoning for categorical syllogisms have been formulated by Guyote and Sternberg (1981) and by Johnson-Laird and Steedman (1978). These models use the idea that the information in premises is represented in the form of examples; for example, "Some jewels are diamonds" might be represented as a symbol for a jewel that is a diamond and another symbol for a jewel that is not a diamond. A representation is formed based on the premises, and is used to evaluate the conclusion. Errors occur because the representations are incomplete; the examples generated by the system often fail to exhaust the possibilities, leading to incorrect conclusions.

V.8. Linear Syllogisms

In a linear syllogism, premises specify ordered relations between pairs of objects, and questions are asked about pairs for which the order was not specified. An example (Egan & Grimes-Farrow, 1982) is:

Circle is darker than square.

Square is darker than triangle.

Is triangle darker than circle?

(An alternative is to ask "Which is darkest?" or "Which is lightest?")

Problems are presented with relations expressed differently, such as "Triangle is lighter than square," or "Triangle is not as dark as square," with the premise information given in different orders, and with different questions.

To answer the question, the information in the premises must be encoded in some representation that enables the answer to be derived. Three hypotheses about representation have been considered.

According to a spatial hypothesis (DeSoto, London & Handel, 1965, Huttenlocher, 1968) information in the premises is integrated into an ordered list, possibly using an image in which symbols are spatially aligned. A representation for the example would be an ordering with circle first, square second, and triangle third, perhaps imagined in a vertical line with the circle at the top. Then a question such as "Is circle darker than triangle?" would be answered by comparing the positions of the circle and the triangle in the ordered representation.

A second hypothesis (Clark, 1969) is that the representation consists of propositions in which individual objects are associated with values of attributes. For the example, circle would be associated with a large degree of darkness, square with a medium degree, and triangle with a small degree. A question would be answered by retrieving representations of the objects in the question and comparing the properties associated with them.

The third hypothesis is that representations of binary relations are stored in memory. This hypothesis assumes the simplest process of representation, since information in memory corresponds directly to the information in the premises. To answer a question, however, a sequence of propositions has to be retrieved; for example to answer "Is circle darker than triangle," both "Circle darker than square" and "Square darker than triangle" have to be retrieved.

The hypothesis that binary relations are represented is ruled out by data obtained by Potts (1974), who had subjects study paragraphs containing series with six terms and asked questions involving pairs that varied in their separation; with the ordering $A > B > C > D > E > F$, $C > D?$ has separation 0, $B > D?$ has separation 1, $B > E?$ has separation 2, and so on. If binary relations are in memory, questions with greater separation should take longer, since answers to these questions require more inferential steps. The finding was the opposite: items with greater separation required less time to respond. This finding has also been obtained with comparisons involving general knowledge, such as the relative sizes of animals (Banks, 1977).

The question whether premises are represented by an integrated spatial array or by propositions associating properties with individual objects has been harder to resolve. Huttenlocher (1968) provided an argument for the spatial hypothesis, including the finding that latency is shorter when the second premise has the third individual as the subject of the sentence (e.g., $A > B$, $C < B$ rather than $A > B$, $B < C$). The interpretation is that the subject imagines placement of the new object in a spatial array, and this is easier if the object is mentioned as the sentence subject than the sentence object. Clark (1969) argued for a propositional representation, presenting evidence that performance is influenced by linguistic factors such as the congruence of questions with premises (e.g., " $A > B$, which is greater?" is easier than " $B < A$, which is greater?").

Sternberg (1980) formulated models that specify stages of processing based on assumptions of a spatial or a propositional representation of premises. He also formulated a model that combines the assumptions, with linguistic factors influencing an initial encoding of premises and relations among propositions influencing a process of converting the information into an integrated spatial array. The combined linguistic-spatial model provided a more accurate account of latency data than either of the simpler models based on linguistic or spatial factors.

Several investigators have provided evidence that linear syllogisms are not solved in a single way by all subjects; rather, different representations are used by different individuals (Mayer, 1979; Sternberg & Weill, 1980). Egan and Grimes-Farrow's (1982) evidence was particularly direct. They used retrospective protocols obtained after solutions of individual problems. The protocols indicated that some subjects used spatial representations consistently, and other subjects sometimes formed representations with individual objects in the problem associated with differing quantitative values of attributes. The protocol evidence was substantiated by analyses showing different influences on subjects' performance depending on the representations they reported using. The order in which objects were mentioned was significant for subjects who used spatial representations, and the linguistic factor of consistency of the

relational term used was significant for subjects who sometimes used individual object propositions.

V.C. Conclusions

Until quite recently there has been little relation between the research on reasoning and research on problem solving of the sorts discussed in the previous sections of this chapter. Sometimes this separation has been justified on the grounds that syllogistic reasoning is "deductive" while problem solving is "inductive." We have seen that this distinction does not hold water. While a syllogism is a deductive structure, finding valid steps or testing whether proposed steps are valid is not a deductive process. Indeed, the major process in evaluation of a propositional or categorical syllogism is an attempt to find a proof of the conclusion, the process that we discuss in Section II.A as the prototypical example of goal-based problem solving. For linear syllogism problems, the major process is an example of inductive problem solving as that concept is used in Section IV, in which the subject forms an integrated representation of the premises using the structure of an ordered list induced from the order relations that the premises state.

From the fact that all reasoning involves problem solving, however, it does not follow that there is no need for special theory in the domain of syllogistic reasoning. To understand human reasoning, we must understand the meanings that people attach to words and the rules of inference that constitute their systems of "natural logic" as well as the structure of the control system that guides their problem solving search. Recent investigations have progressed significantly on these questions.

VI. Conclusions

The literature reviewed in this chapter includes analyses of problem solving in a few dozen tasks. Important general characteristics have emerged in these analyses. One way to express these characteristics is to consider the task of analyzing problem solving in a new domain. Analyses that have been provided give quite strong guidance about the kinds of processes and knowledge structures that one should look for in an investigation of problem solving.

First, it is important to investigate the subjects' knowledge and processes for representing the problem. If the subjects do not have special training in the problem domain, they must construct a problem space that includes representations of the problem materials, the goal, operators, and constraints. If subjects have special training or experience in the domain, their prior knowledge includes general characteristics of the problem space, and their representations of individual problems are based on this general knowledge. Experts are cognizant of general methods that can be used for solving problems, and their representations of problems include use of problem information relevant to the choice of a solution method.

A second major task is to characterize the problem representations that subjects form in their understanding of the problem. In relatively unfamiliar domains, problem solving is primarily a process of search, and the problem representation determines the space of possibilities in which the search will occur. Some basic features of the problem space depend on the problem itself, of course. A problem may present constraints mainly on the operators that are permitted in trying to achieve a goal, or on the arrangement of materials that is acceptable as a solution, or may present materials and require induction of a pattern or rule. These alternatives lead to differences in the problem space: a space of possible sequences of actions, a space of possible solution arrangements, a space of possible structures, or some combination of these.

The problem space constructed by an individual subject also is determined by the method of search that the subject uses, features of the problem that are used, and general knowledge that is applied. In a problem of transforming a situation by a sequence of actions, subjects typically use some form of means-ends analysis. They may distinguish between features of the situation that are more or less essential for the solution, and organize their search by a process of planning that focuses on the more essential features. Searching in a space of possible solution arrangements typically involves generating partial solutions on a trial basis, and is influenced by the subjects' knowledge of constraints that can be used to limit the candidate arrangements that are considered. Solution of induction problems is similarly influenced by the subjects' knowledge of general constraints on possible solutions, which may be used in generating and testing hypotheses or in a process of synthesizing or abstracting structures from the features of individual objects that are provided.

In problem solving for which subjects have special training or experience the problem space of operators and constraints is provided by the subjects' existing knowledge. Knowledge of experts is highly organized, and includes solution methods and concepts for representing

problems at varying degrees of generality and abstraction. For relatively simple problems, experts' knowledge often provides a basis for immediate recognition of solution methods as well as of detailed features relevant to the solution. Their knowledge of relations among methods and operators and of constraints in the domain enables problem-solving performance to occur in a highly organized planful manner.

While the study of problem solving and reasoning has progressed rapidly and achieved a substantial level of knowledge and theory, several significant questions remain largely unanswered. We will mention four of these.

First, while performance of experts on relatively simple problems is beginning to be understood, little is known about their performance on problems that are difficult and deep. It is possible that on problems for which an expert's knowledge does not provide a ready method of solution, the expert resorts to "weak methods" of search and analysis fundamentally similar to those used by novices. It also is possible that powerful processes of reasoning in a domain are acquired by experts, and that these are used in solving problems for which specific solution methods have not been worked out and stored in memory.

A second question, closely related to the first, involves the general nature of problem solving in its more powerful and productive forms. We have referred to discussions of productive thinking by Duncker (1935/1945) and Wertheimer (1945/1959) and have noted progress that has been made on some of the issues that they raised. They also raised a critical issue that has not been addressed strongly in recent discussions. This is the process of constructing more powerful representations of problems by analysis of problem components. The initial representation of a problem frequently does not include important relationships that are required for a meaningful solution, but the problem solver is able to construct a reformulation that includes its important structural features.

A third question for which there are promising preliminary results but much more to be done is the question of learning. Analyses of acquisition require understanding of the skills and knowledge that is acquired, and the significant accomplishments in characterizing skill and knowledge in problem solving provide a promising basis for investigation of learning. Recent proposals regarding acquisition of cognitive skill such as those of Anderson (1982), Anzai and Simon (1979), Neches (1981) and Neves (1981) provide significant steps in the analysis of learning processes.

A fourth question involves the theoretical power of general principles in the analysis of problem solving and reasoning. The analyses that are reviewed in this chapter provide detailed hypotheses about performance in specific tasks, and are strongly testable at the level of their assumptions about specific processes. The assumptions made at a more general level are more heuristic. They involve concepts and principles that provide significant guidance in constructing hypotheses about specific cognitive structures and processes, but they rarely constrain those hypotheses in wholly specifiable ways. The question whether complex processes of problem solving and reasoning are constrained by significant underlying formal principles is an open question. Some investigators (Keil, 1981; VanLehn, Brown & Greeno, in press) have urged that research should attempt to

discover general principles with deductive power that would significantly constrain characteristics of process models. Others (Newell & Simon, 1976) have noted that there are good reasons for expecting that complex cognition is constrained by relatively weak structural principles, of the kind that are characteristic of present theoretical analyses.

A review of any significant body of scientific research can be closed with the remark that much has been accomplished, and much more remains to be done. This seems particularly apt for the psychology of problem solving and reasoning. The progress that has been made in the 1960s and 1970s in this domain has been substantial, and concepts and methods are now available that will enable future investigations to address issues of further significance.

References

- Adamson, R. E. (1952). Functional fixedness as related to problem solving: A repetition of three experiments. Journal of Experimental Psychology, 44, 288-291.
- Adelson, B. (1981). Problem solving and the development of abstract categories in programming languages. Memory and Cognition, 9, 422-433.
- Anderson, J. R. (1975). Computer simulation of a language acquisition system: A first report. In R. L. Solso (Ed.), Information processing and cognition: The Loyola symposium. Hillsdale, NJ: Erlbaum.
- Anderson, J. R. (1976). Language, memory, and thought. Hillsdale, NJ: Erlbaum.
- Anderson, J. R. (1977). Computer simulation of a language acquisition system: A second report. In D. LaBerge & S. J. Samuels (Eds.), Perception and comprehension. Hillsdale, NJ: Erlbaum.
- Anderson, J. R. (1982). Acquisition of cognitive skill. Psychological Review, 89(4), 396-406.
- Anderson, J. R. (1983). The architecture of cognition. Cambridge, MA: Harvard University Press.
- Anderson, J. R., Farrell, R., & Sauers, R. (1983). Learning to program in LISP. Pittsburgh: Carnegie-Mellon University.
- Anderson, J. R., Greeno, J. G., Kline, P. J., & Neves, D. M. (1981). Acquisition of problem-solving skill. In J. R. Anderson (Eds.), Cognitive skills and their acquisition. Hillsdale, NJ: Erlbaum.
- Anderson, J. R., Kline, P. J., & Beasley, C. M. (1979). A general learning theory and its application to schema abstraction. In G. H. Bower (Eds.), The psychology of learning and motivation (Vol. 13). New York: Academic Press, 277-318.
- Anzai, Y., & Simon, H. A. (1979). The theory of learning by doing. Psychological Review, 86, 124-140.
- Ashcraft, M. H. & Battaglia, J. (1978). Cognitive arithmetic: Evidence for retrieval and decision processes in mental addition. Journal of Experimental Psychology: Human Learning and Memory, 5, 527-538.
- Ashcraft, M. H. & Stazyk, E. H. (1981). Mental addition: A test of three verification models. Memory and Cognition, 9, 185-196.
- Atwood, M. E., & Polson, P. G. (1976). A process model for water jug problems. Cognitive Psychology, 8, 191-216.

- Banks, W. P. (1977). Encoding and processing of semantic information in comparative judgments. In G. H. Bower (Ed.), The psychology of learning and motivation: Advances in research and theory (Vol. 11). New York: Academic Press.
- Barnes, G. M., & Whitely, S. E. (1981). Problem restructuring processes for ill-structured verbal analogies. Memory and Cognition, 9, 411-421.
- Bartlett, F. C. (1932). Remembering: A study in experimental and social psychology. Cambridge, England: Cambridge University Press.
- Bartlett, F. C. (1958). Thinking. New York: Basic Books.
- Beilin, H. & Horn, R. (1962). Transition probability effects in anagram problem solving. Journal of Experimental Psychology, 63, 514-518.
- Bereiter, C., & Scardamalia, M. (1982). From conversation to compositions: The role of instruction in a developmental process. In R. Glaser (Ed.), Advances in instructional psychology, (Vol.2). Hillsdale, NJ: Erlbaum.
- Birch, H. G., & Rabinowitz, H. S. (1951). The negative effect of previous experience on productive thinking. Journal of Experimental Psychology, 41, 121-125.
- Bloom, B. S., & Broder, L. J. (1950). Problem solving processes of college students. Chicago: University of Chicago Press.
- Bobrow, D. G. (1968). Natural language input for a computer problem-solving system. In M. Minsky (Eds.), Semantic information processing. Cambridge, MA: MIT Press.
- Bourne, L. E., Jr. (1970). Knowing and using concepts. Psychological Review, 77, 546-556.
- Bourne, L. E., Jr. (1974). An inference model of conceptual rule learning. In R. Solso (Ed.), Theories in cognitive psychology. Washington: Erlbaum.
- Bower, G. H., & Trabasso, T. R. (1964). Concept identification. In R. C. Atkinson (Ed.), Studies in mathematical psychology. Stanford, CA: Stanford University Press, 32-94.
- Braine, M. D. S. (1978). On the relation between the natural logic of reasoning and standard logic. Psychological Review, 85, 1-21.
- Briars, D. J., & Larkin, J. H. (in press). An integrated model of skill in solving elementary word problems. Cognition and instruction.

- Brooks, L. (1978). Nonanalytic concept formation and memory for instances. In E. Rosch, & B. B. Lloyd (Eds.), Cognition and categorization. Hillsdale, NJ: Erlbaum.
- Brown, J. S., & Burton, R. B. (1980). Diagnostic models for procedural bugs in basic mathematical skills. Cognitive Science, 4, 379-426.
- Brown, J. S., Burton, R. R., & de Kleer, J. (1983). Pedagogical, natural language and knowledge engineering techniques in SOPHIE I, II, and III. In D. Sleeman, & J. S. Brown (Eds.), Intelligent tutoring systems. New York: Academic Press.
- Brown, J. S., & VanLehn, K. (1980). Repair theory: A generative theory of bugs in procedural skills. Cognitive Science, 4, 379-426.
- Brown, R. (1973). A first language: The early stages. Cambridge, MA: Harvard University Press.
- Brownell, W. A. (1935). Psychological considerations in the learning and teaching of arithmetic. In The teaching of arithmetic: Tenth yearbook of the National Council of Teachers of Mathematics. New York: Columbia University Press.
- Brownell, W. A. (1941). Arithmetic in grades I and II: A critical summary of new and previously reported research. (Duke University Research Studies in Education, No. 6). Durham: Duke University Press.
- Bruner, J. S., Goodnow, J. J., & Austin, G. A. (1956). A study of thinking. New York: Wiley.
- Bundy, A. (1978). Will it reach the top? Prediction in the mechanics world. Artificial Intelligence, 10, 129-146.
- Burke, R. J., Maier, N. R. F., & Hoffman, L. R. (1966). Function of hints in individual problem solving. Journal of American Psychology, 79, 389-399.
- Buss, A. H. (1953). Rigidity as a function of reversal and nonreversal shifts in the learning of successive discriminations. Journal of Experimental Psychology, 45, 75-81.
- Carpenter, J. A., Moore, O. K., Snyder, C. R., & Lisansky, E. S. (1961). Alcohol and higher-order problem solving. Quarterly Journal of Studies on Alcohol, 22, 183-222.
- Carpenter, T. P., & Moser, J. M. (1981). The development of addition and subtraction problem-solving skills. In T. P. Carpenter, J. M. Moser & T. Romberg (Eds.), Addition and subtraction: A cognitive perspective. Hillsdale, NJ: Erlbaum.

- Chase, W. G., & Simon, H. A. (1973). Perception in chess. Cognitive Psychology, 4, 55-81.
- Chi, M. T. H., Feltovich, P., & Glaser, R. (1981). Categorization and representation of physics problems by experts and novices. Cognitive Science, 5, 121-152.
- Ciesielski, V. B., Bennett, J. S., & Cohen, P. R. (1977). Applications-oriented AI research: Medicine. In A. Barr and E. A. Feigenbaum (Eds.), The handbook of artificial intelligence (Vol.2). Stanford, CA: Heuristech Press.
- Clark, H. H. (1969). Linguistic processes in deductive reasoning. Psychological Review, 76, 387-404.
- deGroot, A. D. (1965). Thought and choice in chess. The Hague: Mouton.
- deKleer, J. (1975). Qualitative and quantitative knowledge in classical mechanics. (AI Lab Technical Report No. AI-TR-352). Cambridge: Massachusetts Institute of Technology.
- DeSoto, C. B., London, M. & Handel, S. (1965). Social reasoning and spatial paralogic. Journal of Personality and Social Psychology, 2, 513-521.
- Dodd, D. H., Kinsman, R., Klipp, R., & Bourne, L. E., Jr. (1971). Effects of logic pretraining on conceptual rule learning. Journal of Experimental Psychology, 88, 119-122.
- Dominowski, R. L., & Ekstrand, B. R. (1967). Direct and associative priming in anagram solving. Journal of Experimental Psychology, 74, 84-86.
- Dunker, K. (1945). On problem solving. Psychological Monographs, 58:5 (Whole No. 270). (Original German version, 1935).
- Egan, D. E. & Grimes-Farrow, D. D. (1982). Differences in mental representations spontaneously adopted for reasoning. Memory and Cognition, 10, 297-307.
- Egan, D. E., & Schwartz, B. J. (1979). Chunking in recall of symbolic drawings. Memory and Cognition, 7, 149-158.
- Eisenstadt, M., & Kareev, Y. (1975). Aspects of human problem solving: The use of internal representations. In D. A. Norman & D. E. Rumelhart, (Eds.), Explorations in cognition (308-346). San Francisco: Freeman.
- Erickson, J. (1971). Problem shifts and hypothesis behavior in concept identification. The American Journal of Psychology, 84(1), 101-111.

- Ericsson, K. A., & Simon, H. A. (1980). Verbal reports as data. Psychological Review, 87, 255-251.
- Ernst, G. W., & Newell, A. (1969). GPS: A case study in generality and problem solving. New York: Academic Press.
- Evans, T. G. (1968). A program for the solution of geometric-analogy intelligence test questions. In M. Minsky (Ed.), Semantic information processing. Cambridge, MA: MIT Press, 271-353.
- Feigenbaum, E. A. (1963). The simulation of verbal learning behavior. In E. A. Feigenbaum & J. Feldman (Eds.), Computers and thought. New York: McGraw-Hill.
- Feltovich, P. J. (1981). Knowledge based components of expertise in medical diagnosis (Report PDS-2). University of Pittsburgh, Learning Research and Development Center.
- Flower, L. S., & Hayes, J. R. (1980). The dynamics of composing: Making plans and juggling constraints. In L. W. Gregg, & E. R. Steinberg (Eds.), Cognitive processes in writing. Hillsdale, NJ: Erlbaum.
- Franks, J. J., & Bransford, J. D. (1971). Abstraction of visual patterns. Journal of Experimental Psychology, 90, 65-74.
- Friedland, P. E. (1979). Knowledge-based experiment design in molecular genetics. (Report No. 79-771). Stanford University, Computer Science Dept., Stanford, CA.
- Gavurin, E. I. (1967). Anagram solution and spatial aptitude. Journal of Psychology, 65, 65-68.(a)
- Gelman, R., & Gallistel, C. R. (1978). The child's understanding of number. Cambridge, MA: Harvard University Press.
- Gentzen, G. (1969). Investigations into logical deduction. In M. E. Szabo (Ed. and trans.), The collected papers of Gerhard Gentzen. Amsterdam: North-Holland. (Originally published, 1935.)
- Gerwin, D. (1974). Information processing, data inferences, and scientific generalization, Behavioral Science, 19(5), 314-325.
- Gick, M. L., & Holyoak, K. J. (1980). Analogical problem solving. Cognitive Psychology, 12, 306-355.
- Gick, M. L., & Holyoak, K. J. (1983). Schema induction and analogical transfer. Cognitive Psychology, 15(1), 1-38.
- Glucksberg, S., & Weisberg, R. W. (1966). Verbal behavior and problem solving: Some effects of labeling in a functional fixedness problem. Journal of Experimental Psychology, 71, 659-664.

- Greeno, J. G. (1976). Indefinite goals in well-structured problems. Psychological Review, 83, 479-491.
- Greeno, J. G. (1978). A study of problem solving. In R. Glaser (Ed.), Advances in instructional psychology, (Vol. 1). Hillsdale, NJ: Erlbaum.
- Greeno, J. G. (1983). Forms of understanding in mathematical problem solving. In S. G. Paris, G. M. Olson, & H. W. Stevenson (Eds.), Learning and motivation in the classroom. Hillsdale, NJ: Erlbaum.
- Greeno, J. G., Magone, M. E., & Chaiklin, S. (1979). Theory of constructions and set in problem solving. Memory and Cognition, 7, 445-461.
- Greeno, J. G., Riley, M. S., & Gelman, R. (in press). Conceptual competence and young children's counting. Cognitive Psychology.
- Greeno, J. G., & Scandura, J. M. (1966). All-or-none transfer based on verbally mediated concepts. Journal of Mathematical Psychology, 3, 388-411.
- Gregg, L. W., & Simon, H. A. (1967). Process models and stochastic theories of simple concept formation. Journal of Mathematical Psychology, 4, 246-276.
- Groen, G. J., & Parkman, J. M. (1972). A chronometric analysis of simple addition. Psychological Review, 79, 329-343.
- Groen, G. J., & Resnick, L. B. (1977). Can preschool children invent addition algorithms? Journal of Educational Psychology, 69, 645-652.
- Grudin, J. (1980). Processes in verbal analogy solution. Journal of Experimental Psychology: Human perception and performance, 6, 67-74.
- Guyote, M. J., & Sternberg, R. J. (1981). A transitive-chain theory of syllogistic reasoning. Cognitive Psychology, 13, 461-525.
- Hayes, J. R. The complete problem solver. (1981). Philadelphia: The Franklin Institute Press.
- Hayes, J. R., & Flower, L. S. (1980). Identifying the organization of writing processes. In L. W. Gregg, & E. R. Steinberg (Eds.), Cognitive processes in writing. Hillsdale, NJ: Erlbaum.
- Hayes, J. R., & Simon, H. A. (1974). Understanding problem instructions. In L. W. Gregg (Ed.), Knowledge and cognition. Hillsdale, NJ: Erlbaum.
- Hayes, J. R., & Simon, H. A. (1977). Psychological differences among problem isomorphs. In N. J. Castellan, P. B. Pisoni, & G. R. Potts (Eds.), Cognitive theory (Vol. 2). Hillsdale, NJ: Erlbaum.

- Hayes-Roth, B., & Hayes Roth, F. (1978). Cognitive Processes in Planning. (Report R-2366-JNR). Rand Corporation, Santa Monica, CA.
- Haygood, R. C. & Bourne, L. E., Jr. (1965). Attribute- and rule-learning aspects of conceptual behavior. Psychological Review, 72, 175-195.
- Heller, J. I. (1979). Cognitive processing in verbal analogy solution. Unpublished doctoral dissertation, University of Pittsburgh.
- Henley, N. M. (1969). A psychological study of the semantics of animal terms. Journal of Verbal Learning and Verbal Behavior, 8, 176-184.
- Huesmann, L. R., & Cheng, C. (1973). A theory for the induction of mathematical functions. Psychological Review, 80, 126-138.
- Hull, C. L. (1920). Quantitative aspects of the evolution of concepts. Psychological Monographs, 28 (Whole No. 20).
- Hull C. L. (1930). Knowledge and purpose as as habit mechanisms. Psychological Review, 37, 241-256.
- Hull, C. L. (1943). Principles of behavior: An introduction to behavior theory. New York: Appleton-Century-Crofts.
- Hull, C. L. (1952). A behavior system. New Haven: Yale University Press.
- Hunt, E. B. (1962). Concept learning: An information processing problem. New York: Wiley.
- Hunt, E. B., Marin, J., & Stone, P. I. (1966). Experiments in induction. New York: Academic Press.
- Huttenlocher, J. (1968). Constructing spatial images: A strategy in reasoning. Psychological Review, 75, 550-560.
- Janis, I. L., & Frick, F. (1943). The relationship between attitudes toward conclusions and errors in judging logical validity of syllogisms. Journal of Experimental Psychology, 33, 73-77.
- Jeffries, R., Polson, P. G., Razran, L., & Atwood, M. E. (1977). A process model for missionaries-cannibals and other river crossing problems. Cognitive Psychology, 9, 412-440.
- Jongman, R. W. (1968). Het oog van de Meester. Amsterdam: van Gorcum.
- Johnson, P. E., Duran, A. S., Hassebrock, F., Moller, J., Prietula, M., Feltovich, P. J., & Swanson, D. B. (1981). Expertise and error in diagnostic reasoning. Cognitive Science, 5, 235-283.

- Johnson-Laird, P. N., & Steedman, M. (1978). The psychology of syllogisms. Cognitive Psychology, 10, 64-99.
- Judd, C. M. (1908). The relation of special training to general intelligence. Educational Review, 36, 28-42.
- Katona, G. (1940). Organizing and memorizing. New York: Columbia University Press.
- Keil, F. C. (1981). Constraints on knowledge and cognitive development. Psychological Review, 88, 197-227.
- Kelleher, R. T. (1956). Discrimination learning as a function of reversal and nonreversal shifts. Journal of Experimental Psychology, 51, 379-384.
- Kendler, H. H. & D'Amato, M. F. (1955). A comparison of reversal and nonreversal shifts in human concept formation behavior. Journal of Experimental Psychology, 49, 165-174.
- Kendler, H. H., & Kendler, T. S. (1959). Reversal and nonreversal shifts in kindergarten children. Journal of Experimental Psychology, 58, 56-60.
- Kendler, T.S., & Kendler, H. H. (1960). Learning, development, and thinking. Annals of the New York Academy of Sciences, 91, 52-63.
- Kintsch, W. (1974). The representation of meaning in memory. Hillsdale, NJ: Erlbaum.
- Konler, W. (1929). The mentality of apes. New York: Harcourt Brace.
- Kotovskiy, K., & Simon, H. A. (1973). Empirical tests of a theory of human acquisition of concepts for sequential events. Cognitive Psychology, 4, 399-424.
- Langley, P. (1981). Data-driven discovery of physical laws. Cognitive Science, 5(1), 31-54.
- Langley, P. Bradshaw, G. L., & Simon, H. A. (in press). Rediscovering chemistry with the BACON system. R. Michalski, J. Carbonell, & T. Mitchell (Eds.), Machine Learning: An artificial intelligence approach. Palo Alto, CA: Tioga Press. formulas.
- Larkin, J. H. (1981). Enriching formal knowledge: A model for learning to solve problems in physics. In J. R. Anderson (Ed.), Cognitive skills and their acquisition. Hillsdale, NJ: Erlbaum.
- Larkin, J. H. (1982). Spatial knowledge in solving physics problems. (C.I.P.434). Carnegie-Mellon University, Department of Psychology.
- Larkin, J. H., McDermott, J., Simon, D. P., & Simon, H. A. (1979). Models of competence in solving physics problems (CIP Working Paper No. 408). Pittsburgh, PA: Carnegie-Mellon University.

- Larkin, J. H., & Reif, F. (1979). Understanding and teaching problem-solving in physics. Journal of Science Education, 1(2), 191-203.
- Lefford, A. (1946). The influence of emotional subject matter on logical reasoning. Journal of Genetic Psychology, 34, 127-151.
- Lenat, D. B. (1982). AM: Discovery in mathematics and heuristic search. Stanford, CA: Stanford University.
- Lesgold, A. M., Feltovich, P. J., Glaser, R., & Wang, Y. (1981). The acquisition of perceptual diagnostic skill in radiology (Report PDS-1). University of Pittsburgh, Learning Research and Development Center.
- Levine, M. (1963). Mediating processes in humans at the outset of discrimination learning. Psychological Review, 70, 254-276.
- Lewis, C. H. (1978). Problem system models of practice effects. Unpublished doctoral dissertation, University of Michigan.
- Lindsay, R. K., Buchanan, B. G., & Lederberg, J. (1980). Applications of artificial intelligence for organic chemistry. New York: McGraw Hill.
- Luchins, A. S. (1942). Mechanization in problem solving. Psychological Monographs, 54: 6 (Whole No. 248).
- Magone, M. E. (1977). Goal analysis and feature detection as processes in the solution of an insight problem. Unpublished master's thesis, University of Pittsburgh.
- Maier, N. R. F. (1931). Reasoning and learning. Psychological Review, 38, 332-346.
- Maier, N. R. F. (1945). Reasoning in humans: III. The mechanisms of equivalent stimuli and of reasoning. Journal of Experimental Psychology, 35, 349-360.
- Maltzman, I. (1955). Thinking: From a behavioristic point of view. Psychological Review, 62, 275-286.
- Malin, J. T. (1979). Information-processing load in problem solving by network search. Journal of Experimental Psychology: Human Perception and Performance, 5(2), 379-390.
- Mayer, R. E. (1974). Acquisition processes and resilience under varying testing conditions for structurally different problem-solving procedures. Journal of Educational Psychology, 66, 644-656.
- Mayer, R. E. (1979). Qualitatively different encoding strategies for linear reasoning premises: Evidence for single association and distance theories. Journal of Experimental Psychology: Human Learning and Memory, 5, 1-10.

- McDermott, J., & Larkin, J. H. (1978). Representing textbook physics problems. In Proceedings of the Second National Conference, Canadian Society for Computational Studies of Intelligence. Toronto, Canada: University of Toronto.
- McKeithen, K. B., Reitman, J. R., Ruter, H. H., & Hirtle, S. C. (1981). Knowledge organization and skill differences in computer programmers. Cognitive Psychology, 13, 307-325.
- Millenson, J. R. (1967). Principles of behavioral analysis. New York: MacMillan.
- Miller, G. A., Galanter, E. H., and Pribram, K. H. (1960). Plans and the structure of behavior. New York: Holt, Rinehart & Winston.
- Miller, R. A., Pople, H. E., & Myers, M. D. (1982). Internist-1, An experimental computer-based diagnostic consultant for general medicine. New England Journal of Medicine, 307, 468-476.
- Millward, R. B., & Wickens, T. D. (1974). Concept-identification models. In D. H. Krantz, R. C. Atkinson, R. D. Luce, & P. Suppes (Eds.), Contemporary developments in mathematical psychology (Vol. 1), (pp. 45-100). San Francisco: W. H. Freeman and Company.
- Moeser, S. D., & Bregman, A. S. (1972). The role of reference in the acquisition of a miniature artificial language. Journal of Verbal Learning and Verbal Behavior, 11, 759-769.
- Moore, O. K., & Anderson, S. B. (1954). Modern logic and tasks for experiments on problem solving behavior. Journal of Psychology, 38, 151-160.
- Morgan, J. L., & Newport, E. L. (1981). The role of constituent structure in the induction of artificial language. Journal of Learning and Verbal Behavior, 20, 67-85.
- Mulholland, T. M., Pellegrino, J. W., & Glaser, R. (1980). Components of geometric analogy solution. Cognitive Psychology, 12, 252-284.
- Neches, R. (1981). Models of heuristic procedure modification. Unpublished doctoral dissertation, Carnegie-Mellon University, Pittsburgh.
- Nesher, P. (1982). Levels of description in the analysis of addition and subtraction. In T. P. Carpenter, J. M. Moser, & T. Romberg (Eds.), Addition and subtraction: A cognitive perspective. Hillsdale, NJ: Erlbaum.
- Neumann, P. G. (1974). An attribute frequency model for the abstraction of prototypes. Memory and Cognition, 2, 241-248.

- Neves, D. M. (1981). Learning procedures from examples. Unpublished doctoral dissertation, Carnegie-Mellon University.
- Newell, A., Shaw, J. C., & Simon, H. A. (1957). Preliminary description of the general problem solving program I (GPS I). CIP Working Paper 7. Carnegie-Mellon University, Pittsburgh.
- Newell, A., & Simon, H. A. (1972). Human problem solving. Englewood Cliffs, NJ: Prentice-Hall.
- Newell, A., & Simon, H. A. (1976). Computer science as empirical inquiry: Symbols and search. Communications of the Association for Computing Machinery, (Vol. 19), 11-126.
- Nisbett, R. E., & Wilson, T. D. (1977). Telling more than we know: Verbal reports on mental processes. Psychological Review, 84, 231-259.
- Norman, D. A., & Rumelhart, D. E. (1975). Explorations in cognition. San Francisco: Freeman.
- Novak, G. S. (1976). Computer understanding of physics problems stated in natural language. American Journal of Computational Linguistics, Microfiche 53.
- Osherson, D. N. (1975). Logical abilities in children (Vol. 3). Hillsdale, NJ: Erlbaum.
- Paige, J. M., & Simon, H. A. (1966). Cognitive processes in solving algebra word problems. In B. Kleinmuntz (Ed.), Problem solving. New York: Wiley.
- Pellegrino, J. W. & Glaser, R. (1982). Analyzing aptitudes for learning: Inductive reasoning. In R. Glaser (Ed.), Advances in Instructional Psychology, (Vol. 2), (pp. 269-345). Hillsdale, NJ: Erlbaum.
- Piaget, J. (1952). The child's conception of number. New York: Norton. (Originally published, 1941. In French.)
- Polson, M. C., Restle, F., & Polson, P. G. (1965). Association and discrimination in paired-associates learning. Journal of Experimental Psychology, 69, 47-55.
- Polson, P. G. (1972). A quantitative analysis of the conceptual processes in the Hull paradigm. Journal of Mathematical Psychology, 9, 141-167.
- Polson, P., Atwood, M. E., Jeffries, R., & Turner, A. (1981). The processes involved in designing software. In J. R. Anderson (Ed.), Cognitive skills and their acquisition. Hillsdale, NJ: Erlbaum.

- Pople, H. E. (1982). Heuristic methods for imposing structure on ill-structured problems: the structuring of medical diagnostics. In P. Szolovits (Ed.), Artificial intelligence in medicine. AAAS Symposium Series, Boulder, CO: Westview Press, 119-85.
- Posner, M. I., Goldsmith, R., & Welton, K. E., Jr. (1967). Perceived distance and the classification of distorted patterns. Journal of Experimental Psychology, 73, 28-38.
- Post, E. L. (1943). Formal reductions of the general combinatorial decision problem. American Journal of Math, 65, 197-268.
- Potts, G. R. (1974). Storing and retrieving information about ordered relationships. Journal of Experimental Psychology, 103, 431-439.
- Quillian, M. R. (1968). Semantic memory. In Minsky, M. (Ed.) Semantic information processing. Cambridge, MA: MIT Press, 216-270.
- Reber, A. S. (1967). Implicit learning of artificial grammars. Journal of Verbal Learning and Verbal Behavior, 6, 855-863.
- Reddy, D. R., Erman, L. D., Fennell, R. D., & Neely, R. B. (1973). The HEARSAY speech understanding system: An example of the recognition process. 3rd International Joint Conference on Artificial Intelligence. Stanford, CA.
- Reed, S. K. (1972). Pattern recognition and categorization. Cognitive Psychology, 3, 383-407.
- Reed, S. K., & Ernst, G. W., & Banerji, R. (1974). The role of analogy in transfer between similar problem states. Cognitive Psychology, 6, 435-450.
- Rees, H. J., & Israel, H. C. (1935). An investigation of the establishment and operation of mental sets. Psychological Monographs, 46, Whole No. 210.
- Reif, F., & Heller, J. I. (1981). Knowledge structure and problem solving in physics. (Educational Science Paper 12. Berkeley: University of California, Physics Dept. and Group in Science and Mathematics Education.
- Reitman, J. S. (1976). Skilled perception in GO: Deducing memory structures from inter-response times. Cognitive Psychology, 8, 336-356.
- Reitman, J. S., & Bower, G. H. (1973). Storage and later recognition of exemplars of concepts. Cognitive Psychology, 4, 194-206.
- Reitman, W. R. (1965). Cognition and thought. New York: Wiley.
- Reitman, W. R., & Wilcox, B. (1978). Pattern recognition and pattern-directed inference in a program. In D. A. Waterman and F. Hayes-Roth (Eds.), Pattern-directed inference systems. New York: Academic Press.

- Resnick, L. B. (1983). A developmental theory of number understanding. In H. P. Ginsburg (Ed.), The development of mathematical thinking. New York: Academic Press.
- Restle, F. (1962). The selection of strategies in cue learning. Psychological Review, 69, 329-343.
- Restle, F. (1970). Theory of serial pattern learning: structural trees. Psychological Review, 77, 481-495.
- Riley, M. S., Greeno, J. G., & Heller, J. I. (1983). Development of children's problem-solving ability in arithmetic. In H. P. Ginsburg (Ed.), The development of mathematical thinking. New York: Academic Press.
- Rips, L. J. (1983). Cognitive processes in propositional reasoning. Psychological Review, 90, 38-71.
- Rips, L. J., & Marcus, S. L. (1977). Suppositions and the analysis of conditional sentences. In M. A. Just & P. A. Carpenter (Eds.), Cognitive processes in comprehension. Hillsdale, NJ: Erlbaum.
- Rosch, E. H. (1973). On the internal structure of perceptual and semantic categories. In T. E. Moore (Ed.), Cognitive development and acquisition of language. New York: Academic Press.
- Rosch, E. H. (1975). Cognitive representations of semantic categories. Journal of Experimental Psychology: General, 104, 192-233.
- Rosch, E. H. (1978). Principles of categorization. In E. Rosch, & B. B. Lloyd, (Eds.), Cognition and categorization. Hillsdale, NJ: Erlbaum.
- Rosch, E. H. Mervis, C. B., Gray, W. D., Johnson, D. J., & Boyes-Braem, P. (1976). Basic objects in natural categories. Cognitive Psychology.
- Rumelhart, D. E. and Abrahamson, A. A. (1973). A model for analogical reasoning. Cognitive Psychology, 5, 1-28.
- Sacerdoti, E. D. (1977). A structure for plans and behavior. New York: Elsevier-North Holland.
- Schank, R. C. (1972). Conceptual dependency: A theory of natural language understanding. Cognitive Psychology, 3, 552-631.
- Schank, R. C., & Abelson, R. P. (1978). Scripts, plans, goals, and understanding: An inquiry into human knowledge structures. Hillsdale, NJ: Erlbaum.
- Schoenfeld, A. H. (1979). Explicit heuristic training as a variable in problem-solving performance. Journal of Research in Mathematics Education, 10, 173-187.

- Shortliffe, E. H. (1976). Computer-based medical consultations: Mycin. New York: American Elsevier.
- Silverman, H. (1975). A digitalis therapy advisor. (Rep. MACTR-143). Cambridge: MIT, Computer Science Dept.
- Simon, D. P., & Simon, H. A. (1978). Individual differences in solving physics problems. In R. Siegler (Ed.), Children's thinking: What develops? Hillsdale, NJ: Erlbaum.
- Simon, H. A. (1973). The structure of ill structured problems. Artificial Intelligence, 4, 181-201.
- Simon, H. A., & Barenfeld, M. (1969). Information processing analysis of perceptual processes in problem solving. Psychological Review, 76, 473-483.
- Simon, H. A., & Feigenbaum, E. A. (1964). An information-processing theory of some effects of similarity, familiarization, and meaningfulness in verbal learning. Journal of Verbal Learning on Verbal Behavior, 3, 385-396.
- Simon, H. A., & Gilmarin, K. (1973). A simulation of memory for chess positions. Cognitive Psychology, 2, 29-46.
- Simon, H. A., & Kotovsky, K. (1963). Human acquisition of concepts for sequential patterns. Psychological Review, 70, 534-546.
- Simon, H. A., & Lea, G. (1974). Problem solving and rule induction: A unified vie. In L. W. Gregg (Ed.), Knowledge and cognition. Potomac, MD: Erlbaum.
- Skinner, B. F. (1938). The behavior of organisms: An experimental analysis. New York: Appleton-Century-Crofts.
- Smith, E. E., Shoben, E. J., & Rips, L. J. (1974). Structure and process in semantic memory: A featural model for semantic decisions. Psychological Review, 81(3), 214-241.
- Snow, R. E. (1980). Aptitude processes. In R. E. Snow, P. A. Federico, & W. E. Montague (Eds.), Aptitude, learning, and instruction: Cognitive process analysis analyses of aptitude (Vol. 1). Hillsdale, NJ: Erlbaum.
- Soloway, E., Ehrlich, K., Bonar, J., & Greenspan, J. (1982). What do novices know about programming? In B. Schneiderman & A. Bodre (Eds.), Directions in human-computer interactions. Norwood, NJ: Ablex.
- Staats, A. W. (1966). An integrated-functional learning approach to complex human behavior. In B. Kleinmuntz (Ed.), Problem solving: Research, method, and theory. New York: John Wiley & Sons.

- Stefik, M. (1981). Planning with constraints (MOLGEN: Part 1). Artificial Intelligence, 16, 111-140.
- Sternberg, R. J. (1977). Component processes in analogical reasoning. Psychological Review, 84, 353-378.
- Sternberg, R. J. (1980). Representation and process in linear syllogistic reasoning. Journal of Experimental Psychology: General, 109, 119-159.
- Sternberg, R. J., & Weil, E. M. (1980). An aptitude-strategy interaction in linear syllogistic reasoning. Journal of Educational Psychology, 72, 226-239.
- Sweller, J., & Gee, W. (1978). Einstellung, the sequence effect, and hypothesis theory. Journal of Experimental Psychology: Human Learning and Memory, 4, 513-526.
- Thorndike, E. L. (1923). The psychology of learning. New York: Teachers College, Columbia University.
- Tolman I. (1928). Purposive behavior in animals and men. New York: Century.
- Trabasso, T. R., Rollins, H., & Schaughnessy, E. (1971). Storage and verification states in processing concepts. Cognitive Psychology, 2, 239-289.
- VanLehn, K. (1982). Bugs are not enough: Empirical studies of bugs, impasses, and repairs in procedural skills. Palo Alto: Xerox PARC.
- VanLehn, K. (1983). On the representation of procedures in repair theory. In H. P. Ginsburg (Ed.), The development of mathematical thinking. New York: Academic Press.
- VanLehn, K., Brown, J. S., & Greeno, J. G. (in press). Competitive argumentation in computational theories of cognition. In W. Kintsch, J. Miller, & P. Polson (Eds.), Methods and tactics in cognitive science. New York: Erlbaum.
- Vergnaud, G. (1982). A classification of cognitive tasks and operations of thought involved in addition and subtraction problems. In T. P. Carpenter, J. M. Moser, & T. Romberg (Eds.), Addition and subtraction: A cognitive perspective. Hillsdale, NJ: Erlbaum.
- Voss, J. F., Greene, T. R., Post, T. A., & Penner, B. C. (1983). Problem solving skill in the social sciences. In G. H. Bower (Ed.), The psychology of learning and motivation: Advances in research theory. (Vol. 17), (pp. 165-213). New York: Academic Press.

- Weiss, S., Kulikowski, C., & Safir, A. (1977). A model-based consultation system for the long-term management of glaucoma. IJCAI5, 826-832.
- Weisberg, R.W., & Alba, J. W. (1981). Gestalt theory, insight, and past experience: Reply to Dominowski. Journal of Experimental Psychology: General, 110(2), 193-203.
- Weisberg, R., & Suls, J. M. (1973). An information-processing model of Duncker's candle problem. Cognitive Psychology, 4, 255-276.
- Wertheimer, M. (1959). Productive thinking, enlarged edition. New York: Harper, (first edition published, 1945).
- Wexler, K., & Cullicover, P. W. (1980). Formal principles of language acquisition. Cambridge, MA: The MIT Press.
- Wickens, T. D., & Millward, R. B. (1971). Attribute elimination strategies for concept identification with practiced subjects. Journal of Mathematical Psychology, 8, 453-480.
- Wilkins, D. (1980). Using patterns and plans in chess. Artificial Intelligence, 14, 165-203.
- Wilkins, M. C. (1928). The effect of changed material on ability to do formal syllogistic reasoning. Archives of Psychology, No. 102.
- Williams, G. (1971). A model of memory in concept learning. Cognitive Psychology, 2, 158-184.
- Winkelman, J. H., & Schmidt, J. (1974). Associative confusions in mental arithmetic. Journal of Experimental Psychology, 102, 734-736.
- Winograd, T. (1972). Understanding natural language. Cognitive Psychology, 3, 1-191.
- Woodworth, R. S., & Sells, S. B. (1935). An atmosphere effect in formal syllogistic reasoning. Journal of Experimental Psychology, subtraction. Cognitive Science, 6, 153-178.
- Woodworth, R. S. (1938). Experimental psychology. New York: Holt.
- Young, R. M., & O'Shea, T. (1981). Errors in children's subtraction. Cognitive Science, 6, 153-178.

Navy

- 1 Robert Ahlers
Code N711
Human Factors Laboratory
NAVTRAEQUIPCEN
Orlando, FL 32813
- 1 Dr. Ed Aiken
Navy Personnel R&D Center
San Diego, CA 92152
- 1 Dr. Meryl S. Baker
Navy Personnel R&D Center
San Diego, CA 92152
- 1 Dr. Nick Bond
Office of Naval Research
Liaison Office, Far East
APO San Francisco, CA 96503
- 1 Dr. Richard Cantone
Navy Research Laboratory
Code 7510
Washington, DC 20375
- 1 Dr. Fred Chang
Navy Personnel R&D Center
San Diego, CA 92152
- 1 Dr. Stanley Collyer
Office of Naval Technology
800 N. Quincy Street
Arlington, VA 22217
- 1 CDR Mike Curran
Office of Naval Research
800 N. Quincy St.
Code 270
Arlington, VA 22217
- 1 DR. PAT FEDERICO
Code P13
NPRDC
San Diego, CA 92152
- 1 Dr. Jude Franklin
Code 7510
Navy Research Laboratory
Washington, DC 20375
- 1 Dr. Jim Hollan
Code 14
Navy Personnel R & D Center
San Diego, CA 92152

Navy

- 1 Dr. Ed Hutchins
Navy Personnel R&D Center
San Diego, CA 92152
- 1 Dr. Norman J. Kerr
Chief of Naval Technical Training
Naval Air Station Memphis (75)
Millington, TN 38054
- 1 Dr. Peter Kincaid
Training Analysis & Evaluation Group
Dept. of the Navy
Orlando, FL 32813
- 1 Dr. William L. Maloy (02)
Chief of Naval Education and Training
Naval Air Station
Pensacola, FL 32508
- 1 Dr. Joe McLachlan
Navy Personnel R&D Center
San Diego, CA 92152
- 1 Dr. James McMichael
Navy Personnel R&D Center
San Diego, CA 92152
- 1 Dr William Montague
NPRDC Code 13
San Diego, CA 92152
- 1 Library, Code P201L
Navy Personnel R&D Center
San Diego, CA 92152
- 1 Technical Director
Navy Personnel R&D Center
San Diego, CA 92152
- 6 Commanding Officer
Naval Research Laboratory
Code 2627
Washington, DC 20390
- 1 Office of Naval Research
Code 433
800 N. Quincy Street
Arlington, VA 22217
- 6 Personnel & Training Research Group
Code 442PT
Office of Naval Research
Arlington, VA 22217

Navy

- 1 Psychologist
ONR Branch Office
1030 East Green Street
Pasadena, CA 91101
- 1 Dr. Robert G. Smith
Office of Chief of Naval Operations
OP-987H
Washington, DC 20350
- 1 Dr. Alfred F. Smode, Director
Training Analysis & Evaluation Group
Dept. of the Navy
Orlando, FL 32813
- 1 Dr. Richard Snow
Liaison Scientist
Office of Naval Research
Branch Office, London
Box 39
FPO New York, NY 09510
- 1 Dr. Richard Sorensen
Navy Personnel R&D Center
San Diego, CA 92152
- 1 Dr. Frederick Steinheiser
CNO - OP115
Navy Annex
Arlington, VA 20370
- 1 Dr. Thomas Sticht
Navy Personnel R&D Center
San Diego, CA 92152
- 1 Roger Weissinger-Baylon
Department of Administrative Sciences
Naval Postgraduate School
Monterey, CA 93940
- 1 Mr John H. Wolfe
Navy Personnel R&D Center
San Diego, CA 92152
- 1 Dr. Wallace Wulfeck, III
Navy Personnel R&D Center
San Diego, CA 92152

Marine Corps

- 1 H. William Greenup
Education Advisor (E031)
Education Center, MCDEC
Quantico, VA 22134
- 1 Special Assistant for Marine
Corps Matters
Code 100M
Office of Naval Research
800 N. Quincy St.
Arlington, VA 22217
- 1 DR. A.L. SLAFKOSKY
SCIENTIFIC ADVISOR (CODE RD-1)
HQ, U.S. MARINE CORPS
WASHINGTON, DC 20380

Army

- 1 Technical Director
U. S. Army Research Institute for the
Behavioral and Social Sciences
5001 Eisenhower Avenue
Alexandria, VA 22333
- 1 Dr. Beatrice J. Farr
U. S. Army Research Institute
5001 Eisenhower Avenue
Alexandria, VA 22333
- 1 Dr. Harold F. O'Neil, Jr.
Director, Training Research Lab
Army Research Institute
5001 Eisenhower Avenue
Alexandria, VA 22333
- 1 Commander, U.S. Army Research Institute
for the Behavioral & Social Sciences
ATTN: PERI-BR (Dr. Judith Orasanu)
5001 Eisenhower Avenue
Alexandria, VA 22333
- 1 Joseph Psotka, Ph.D.
ATTN: PERI-1C
Army Research Institute
5001 Eisenhower Ave.
Alexandria, VA 22333
- 1 Dr. Robert Sasmor
U. S. Army Research Institute for the
Behavioral and Social Sciences
5001 Eisenhower Avenue
Alexandria, VA 22333

Air Force

- 1 Dr. Earl A. Alluisi
HQ, AFHRL (AFSC)
Brooks AFB, TX 78235
- 1 Mr. Raymond E. Christal
AFHRL/MOE
Brooks AFB, TX 78235
- 1 Bryan Dallman
AFHRL/LRT
Lowry AFB, CO 80230
- 1 Dr. Genevieve Haddad
Program Manager
Life Sciences Directorate
AFOSR
Bolling AFB, DC 20332
- 1 Dr. T. M. Longridge
AFHRL/OTE
Williams AFB, AZ 85224
- 1 Dr. John Tangney
AFOSR/NL
Bolling AFB, DC 20332
- 1 Dr. Joseph Yasatuke
AFHRL/LRT
Lowry AFB, CO 80230

Department of Defense

- 12 Defense Technical Information Center
Cameron Station, Bldg 5
Alexandria, VA 22314
Attn: TC
- 1 Military Assistant for Training and
Personnel Technology
Office of the Under Secretary of Defense
for Research & Engineering
Room 3D129, The Pentagon
Washington, DC 20301
- 1 Major Jack Thorpe
DARPA
1400 Wilson Blvd.
Arlington, VA 22209
- 1 Dr. Robert A. Wisher
OUSDRE (ELS)
The Pentagon, Room 3D129
Washington, DC 20301

Civilian Agencies

- 1 Dr. Patricia A. Butler
NIE-BRN Bldg, Stop # 7
1200 19th St., NW
Washington, DC 20208
- 1 Dr. Paul G. Chapin
Linguistics Program
National Science Foundation
Washington, DC 20550
- 1 Dr. Susan Chipman
Learning and Development
National Institute of Education
1200 19th Street NW
Washington, DC 20208
- 1 Dr. Arthur Melmed
724 Brown
U. S. Dept. of Education
Washington, DC 20208
- 1 Dr. Andrew R. Molnar
Office of Scientific and Engineering
Personnel and Education
National Science Foundation
Washington, DC 20550
- 1 Dr. Ramsay W. Selden
National Institute of Education
1200 19th St., NW
Washington, DC 20208
- 1 Chief, Psychological Research Branch
U. S. Coast Guard (G-P-1/2/TP42)
Washington, DC 20593
- 1 Dr. Frank Withrow
U. S. Office of Education
400 Maryland Ave. SW
Washington, DC 20202
- 1 Dr. Joseph L. Young, Director
Memory & Cognitive Processes
National Science Foundation
Washington, DC 20550

Private Sector

- 1 Dr. John R. Anderson
Department of Psychology
Carnegie-Mellon University
Pittsburgh, PA 15213
- 1 Dr. Alan Baddeley
Medical Research Council
Applied Psychology Unit
15 Chaucer Road
Cambridge CB2 2EF
ENGLAND
- 1 Eva L. Baker
Director
UCLA Center for the Study of Evaluation
145 Moore Hall
University of California, Los Angeles
Los Angeles, CA 90024
- 1 Mr. Avron Barr
Department of Computer Science
Stanford University
Stanford, CA 94305
- 1 Dr. Menucha Birenbaum
School of Education
Tel Aviv University
Tel Aviv, Ramat Aviv 69978
Israel
- 1 Dr. John Black
Yale University
Box 11A, Yale Station
New Haven, CT 06520
- 1 Dr. John S. Brown
XEROX Palo Alto Research Center
3333 Coyote Road
Palo Alto, CA 94304
- 1 Dr. Glenn Bryan
6208 Poe Road
Bethesda, MD 20817
- 1 Dr. Bruce Buchanan
Department of Computer Science
Stanford University
Stanford, CA 94305
- 1 Dr. Jaime Carbonell
Carnegie-Mellon University
Department of Psychology
Pittsburgh, PA 15213

Private Sector

- 1 Dr. Pat Carpenter
Department of Psychology
Carnegie-Mellon University
Pittsburgh, PA 15213
- 1 Dr. Micheline Chi
Learning R & D Center
University of Pittsburgh
3939 O'Hara Street
Pittsburgh, PA 15213
- 1 Dr. William Clancey
Department of Computer Science
Stanford University
Stanford, CA 94306
- 1 Dr. Michael Cole
University of California
at San Diego
Laboratory of Comparative
Human Cognition - D003A
La Jolla, CA 92093
- 1 Dr. Allan M. Collins
Bolt Beranek & Newman, Inc.
50 Moulton Street
Cambridge, MA 02138
- 1 Dr. Lynn A. Cooper
LRDC
University of Pittsburgh
3939 O'Hara Street
Pittsburgh, PA 15213
- 1 Dr. Emmanuel Donchin
Department of Psychology
University of Illinois
Champaign, IL 61820
- 1 Dr. Thomas M. Duffy
Department of English
Carnegie-Mellon University
Schenley Park
Pittsburgh, CA 15213
- 1 Dr. Anders Ericsson
Department of Psychology
University of Colorado
Boulder, CO 80309

Private Sector

- 1 Dr. Paul Feltovich
Department of Medical Education
Southern Illinois University
School of Medicine
P.O. Box 3926
Springfield, IL 62708
- 1 Professor Reuven Feuerstein
HWCRI Rehov Karmon 6
Bet Hakerem
Jerusalem
Israel
- 1 Mr. Wallace Feurzeig
Department of Educational Technology
Bolt Beranek & Newman
10 Moulton St.
Cambridge, MA 02238
- 1 Dr. Dexter Fletcher
University of Oregon
Department of Computer Science
Eugene, OR 97403
- 1 Dr. John R. Frederiksen
Bolt Beranek & Newman
50 Moulton Street
Cambridge, MA 02138
- 1 Dr. Michael Genesereth
Department of Computer Science
Stanford University
Stanford, CA 94305
- 1 Dr. Don Gentner
Center for Human Information Processing
University of California, San Diego
La Jolla, CA 92093
- 1 Dr. Dedre Gentner
Bolt Beranek & Newman
10 Moulton St.
Cambridge, MA 02138
- 1 Dr. Robert Glaser
Learning Research & Development Center
University of Pittsburgh
3939 O'Hara Street
PITTSBURGH, PA 15260
- 1 Dr. Marvin D. Glock
217 Stone Hall
Cornell University
Ithaca, NY 14853

Private Sector

- 1 Dr. Josph Goguen
SRI International
333 Ravenswood Avenue
Menlo Park, CA 94025
- 1 Dr. Daniel Gopher
Faculty of Industrial Engineering
& Management
TECHNION
Haifa 32000
ISRAEL
- 1 Dr. Bert Green
Johns Hopkins University
Department of Psychology
Charles & 34th Street
Baltimore, MD 21218
- 1 DR. JAMES G. GREENO
LRDC
UNIVERSITY OF PITTSBURGH
3939 O'HARA STREET
PITTSBURGH, PA 15213
- 1 Dr. Barbara Hayes-Roth
Department of Computer Science
Stanford University
Stanford, CA 95305
- 1 Dr. Frederick Hayes-Roth
Teknowledge
525 University Ave.
Palo Alto, CA 94301
- 1 Dr. Joan I. Heller
Graduate Group in Science and
Mathematics Education
c/o School of Education
University of California
Berkeley, CA 94720
- 1 Dr. James R. Hoffman
Department of Psychol.
University of Delaware
Newark, DE 19711
- 1 American Institutes for Research
1055 Thomas Jefferson St., N.W.
Washington, DC 20007
- 1 Glenda Greenwald, Ed.
Human Intelligence Newsletter
P. O. Box 1163
Birmingham, MI 48012

Private Sector

- 1 Dr. Earl Hunt
Dept. of Psychology
University of Washington
Seattle, WA 98105
- 1 Robin Jeffries
Computer Research Center
Hewlett-Packard Laboratories
1501 Page Mill Road
Palo Alto, CA 94304
- 1 Dr. Marcel Just
Department of Psychology
Carnegie-Mellon University
Pittsburgh, PA 15213
- 1 Dr. David Kieras
Department of Psychology
University of Arizona
Tucson, AZ 85721
- 1 Dr. Walter Kintsch
Department of Psychology
University of Colorado
Boulder, CO 80302
- 1 Dr. Stephen Kosslyn
1236 William James Hall
33 Kirkland St.
Cambridge, MA 02138
- 1 Dr. Pat Langley
The Robotics Institute
Carnegie-Mellon University
Pittsburgh, PA 15213
- 1 Dr. Jill Larkin
Department of Psychology
Carnegie Mellon University
Pittsburgh, PA 15213
- 1 Dr. Alan Lesgold
Learning R&D Center
University of Pittsburgh
3939 O'Hara Street
Pittsburgh, PA 15260
- 1 Dr. Jim Levin
University of California
at San Diego
Laboratory of Comparative
Human Cognition - D003A
La Jolla, CA 92093

Private Sector

- 1 Dr. Michael Levine
Department of Educational Psychology
210 Education Bldg.
University of Illinois
Champaign, IL 61801
- 1 Dr. Marcia C. Linn
Lawrence Hall of Science
University of California
Berkeley, CA 94720
- 1 Dr. Don Lyon
AFHRL/OT (UDRI)
Williams AFB, AZ 85225
- 1 Dr. Jay McClelland
Department of Psychology
MIT
Cambridge, MA 02139
- 1 Dr. James R. Miller
Computer*Thought Corporation
1721 West Plano Highway
Plano, TX 75075
- 1 Dr. Mark Miller
Computer*Thought Corporation
1721 West Plano Parkway
Plano, TX 75075
- 1 Dr. Tom Moran
Xerox PARC
3333 Coyote Hill Road
Palo Alto, CA 94304
- 1 Dr. Allen Munro
Behavioral Technology Laboratories
1845 Elena Ave., Fourth Floor
Redondo Beach, CA 90277
- 1 Dr. Donald A Norman
Cognitive Science, C-015
Univ. of California, San Diego
La Jolla, CA 92093
- 1 Dr. Jesse Orlansky
Institute for Defense Analyses
1801 N. Beauregard St.
Alexandria, VA 22311
- 1 Prof. Seymour Papert
20C-109
Massachusetts Institute of Technology
Cambridge, MA 02139

Private Sector

- 1 Dr. James W. Pellegrino
University of California,
Santa Barbara
Dept. of Psychology
Santa Barabara, CA 93106
- 1 Dr. Nancy Pennington
University of Chicago
Graduate School of Business
1101 E. 58th St.
Chicago, IL 60637
- 1 DR. PETER POLSON
DEPT. OF PSYCHOLOGY
UNIVERSITY OF COLORADO
BOULDER, CO 80309
- 1 Dr. Mike Posner
Department of Psychology
University of Oregon
Eugene, OR 97403
- 1 Dr. Lynn Reder
Department of Psychology
Carnegie-Mellon University
Schenley Park
Pittsburgh, PA 15213
- 1 Dr. Fred Reif
Physics Department
University of California
Berkeley, CA 94720
- 1 Dr. Lauren Resnick
LRDC
University of Pittsburgh
3939 O'Hara Street
Pittsburgh, PA 1521
- 1 Dr. Jeff Richardson
Denver Research Institute
University of Denver
Denver, CO 80208
- 1 Mary S. Riley
Program in Cognitive Science
Center for Human Information Processing
University of California, San Diego
La Jolla, CA 92093
- 1 Dr. Andrew M. Rose
American Institutes for Research
1055 Thomas Jefferson St. NW
Washington, DC 20007

Private Sector

- 1 Dr. Ernst Z. Rothkopf
Bell Laboratories
Murray Hill, NJ 07974
- 1 Dr. William B. Rouse
Georgia Institute of Technology
School of Industrial & Systems
Engineering
Atlanta, GA 30332
- 1 Dr. David Rumelhart
Center for Human Information Processing
Univ. of California, San Diego
La Jolla, CA 92093
- 1 Dr. Michael J. Samet
Perceptronics, Inc
6271 Variel Avenue
Woodland Hills, CA 91364
- 1 Dr. Roger Schank
Yale University
Department of Computer Science
P.O. Box 2158
New Haven, CT 06520
- 1 Dr. Walter Schneider
Psychology Department
603 E. Daniel
Champaign, IL 61820
- 1 Dr. Alan Schoenfeld
Mathematics and Education
The University of Rochester
Rochester, NY 14627
- 1 Mr. Colin Sheppard
Applied Psychology Unit
Admiralty Marine Technology Est.
Teddington, Middlesex
United Kingdom
- 1 Dr. H. Wallace Sinaiko
Program Director
Manpower Research and Advisory Services
Smithsonian Institution
801 North Pitt Street
Alexandria, VA 22314
- 1 Dr. Edward E. Smith
Bolt Beranek & Newman, Inc.
50 Moulton Street
Cambridge, MA 02138

Private Sector

- 1 Dr. Elliott Soloway
Yale University
Department of Computer Science
P.O. Box 2158
New Haven, CT 06520
- 1 Dr. Kathryn T. Spoehr
Psychology Department
Brown University
Providence, RI 02912
- 1 Dr. Robert Sternberg
Dept. of Psychology
Yale University
Box 11A, Yale Station
New Haven, CT 06520
- 1 Dr. Albert Stevens
Bolt Beranek & Newman, Inc.
10 Moulton St.
Cambridge, MA 02238
- 1 David E. Stone, Ph.D.
Hazeltine Corporation
7680 Old Springhouse Road
McLean, VA 22102
- 1 DR. PATRICK SUPPES
INSTITUTE FOR MATHEMATICAL STUDIES IN
THE SOCIAL SCIENCES
STANFORD UNIVERSITY
STANFORD, CA 94305
- 1 Dr. Kikumi Tatsuoka
Computer Based Education Research Lab
252 Engineering Research Laboratory
Urbana, IL 61801
- 1 Dr. Maurice Tatsuoka
220 Education Bldg
1310 S. Sixth St.
Champaign, IL 61820
- 1 Dr. Perry W. Thorndyke
Perceptronics, Inc.
545 Middlefield Road, Suite 140
Menlo Park, CA 94025
- 1 Dr. Douglas Towne
Univ. of So. California
Behavioral Technology Labs
1845 S. Elena Ave.
Redondo Beach, CA 90277

Private Sector

- 1 Dr. Kurt Van Lehn
Xerox PARC
3333 Coyote Hill Road
Palo Alto, CA 94304
- 1 Dr. Keith T. Wescourt
Perceptronics, Inc.
545 Middlefield Road, Suite 140
Menlo Park, CA 94025
- 1 William B. Whitten
Bell Laboratories
2D-610
Holmdel, NJ 07733
- 1 Dr. Thomas Wickens
Department of Psychology
Franz Hall
University of California
405 Hilgarde Avenue
Los Angeles, CA 90024
- 1 Dr. Mike Williams
IntelliGenetics
124 University Avenue
Palo Alto, CA 94301
- 1 Dr. Joseph Wohl
Alphatech, Inc.
2 Burlington Executive Center
111 Middlesex Turnpike
Burlington, MA 01803

8
DTIC